

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 ; Search time 54.03 Seconds
(without alignments)
4889.196 Million cell updates/sec

Title: US-09-695-795-4
Perfect score: 7883
Sequence: 1 MSIRLPHSIDRSASRKSQSHL.....EGSDIALDSATDFAVSPGP 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

vs 10/23/99

Result No.	Score	Query Match	Length	DB	ID	Description
1	7883	100.0	1527	11	Q9ES67	Q9ES67 ratu
2	6482.5	82.2	1522	4	O15085	O15085 homo sapien
3	1822.5	23.1	1544	4	O9NZNS	O9NZNS homo sapien
4	1223	15.5	920	11	O9IVL3	O9IVL3 mus musculu
5	1216.5	15.4	919	11	O61210	O61210 mus musculu
6	1196.5	15.2	912	4	O92888	O92888 homo sapien
7	1180	15.0	919	11	O9Z116	O9Z116 ratu
8	1151	14.6	853	4	O9BSB1	O9BSB1 homo sapien
9	1139.5	14.5	869	4	O00513	O00513 homo sapien
10	1139	14.4	879	4	O96F17	O96F17 homo sapien
11	1032	13.0	750	4	O15086	O15086 homo sapien
12	962.5	12.2	540	4	O96BF4	O96BF4 homo sapien
13	944.5	12.0	2559	5	O44381	O44381 drosophila
14	936.5	11.9	2559	5	O9V7X3	O9V7X3 drosophila
15	922.5	11.7	2559	5	O44113	O44113 drosophila
16	672	8.5	979	5	O19402	O19402 caenorhabd1

17	556.5	7.1	2817	4	Q96P79	Q96P79 homo sapien
18	547.5	6.9	2813	4	O96JPE	O96JPE homo sapien
19	531	6.7	1050	4	O60274	O60274 homo sapien
20	525.5	6.7	958	4	O9H023	O9H023 homo sapien
21	517.5	6.6	1429	4	O9Y5T6	O9Y5T6 homo sapien
22	492.5	6.2	985	11	Q9ESG7	Q9ESG7 mus musculu
23	486.5	6.2	985	11	O9Z3E0	O9Z3E0 mus musculu
24	457	5.8	893	4	O9Y5T0	O9Y5T0 homo sapien
25	451	5.7	930	4	O96BS1	O96BS1 homo sapien
26	449.5	5.7	164	11	O98074	O98074 mus musculu
27	442.5	5.6	864	4	O94827	O94827 homo sapien
28	346	4.4	1309	5	O9V1V0	O9V1V0 drosophila
29	343	4.4	1721	4	O95216	O95216 homo sapien
30	340	4.3	1714	11	O9Z0R4	O9Z0R4 mus musculu
31	340	4.3	1721	4	O9UNK2	O9UNK2 homo sapien
32	328	4.2	1676	4	O9UIG4	O9UIG4 homo sapien
33	328	4.2	1681	4	O9NYG0	O9NYG0 homo sapien
34	326	4.1	1686	4	O9NZM3	O9NZM3 homo sapien
35	310.5	3.9	790	4	O9NVK9	O9NVK9 homo sapien
36	301.5	3.8	1658	11	O9Z0R6	O9Z0R6 mus musculu
37	297.5	3.8	1737	5	O9TW28	O9TW28 dicystostell
38	297	3.8	1804	11	O9Z1Q1	O9Z1Q1 mesocricetu
39	296.5	3.8	1821	11	O9R0C4	O9R0C4 mus musculu
40	293.5	3.7	1234	5	O9VZ85	O9VZ85 drosophila
41	292.5	3.7	841	4	O94989	O94989 homo sapien
42	291.5	3.7	660	5	O9NZV4	O9NZV4 caenorhabd1
43	288.5	3.7	1065	11	O91VS8	O91VS8 mus musculu
44	288	3.7	2137	4	O15021	O15021 homo sapien
45	279.5	3.5	522	4	O9H8F4	O9H8F4 homo sapien

ALIGNMENTS

RESULT 1
ID Q9ES67 PRELIMINARY; PRT; 1527 AA.
AC Q9ES67;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RHOGEF GLUTAMATE TRANSPORT MODULATOR GTRAP48.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2113160; Pubmed=11242047;
RA Jackson M., Song W., Liu M.Y., Jin L., Dykes-Hoberg M., Lin C.-L.G.,
RA Bowers W.J., Federoff H.J., Sternweis P.C., Rothstein J.D.;
RT "Modulation of the neuronal glutamate transporter EAAT4 by two
RT interacting proteins";
RL Nature 410:89-93(2001).
CC -
CC - SIMILARITY: CONTAINS 1 RGS DOMAIN.
DR EMBL; AF225961; AAC28597.1; -
DR HSSP; Q12923; 3PDZ.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000342; RGS.
DR InterPro: IPR000219; RHOGEF.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00621; RHOGEF; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00315; RGS; 1.
DR SMART; SM00325; RHOGEF; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50132; RGS; 1.
SQ SEQUENCE 1527 AA; 168533 MW; ABAEA20F41A3A9A CRC64;

Query Match 100.0%; Score 7883; DB 11; Length 1527;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLPHSIDRSASKKOSHLSPISASMLSSLSLGDSTPERTSPSHRQPSDSETTAGL 60
 DB 1 MSRLPHSIDRSASKKOSHLSPISASMLSSLSLGDSTPERTSPSHRQPSDSETTAGL 60

QY 61 VORCVIIOKDQHGFGFTVSQDRIVLYQSVRPGGAAMKAGYEGEDRIIKVNGTWTNSSHL 120
 DB 61 VORCVIIOKDQHGFGFTVSQDRIVLYQSVRPGGAAMKAGYEGEDRIIKVNGTWTNSSHL 120

QY 121 EYVKLKSGVAYALTLTGSSPSPSVGSLQONSVAGVLRVNPITPPPPPPPLPPPH 180
 DB 121 EYVKLKSGVAYALTLTGSSPSPSVGSLQONSVAGVLRVNPITPPPPPPPLPPPH 180

QY 181 TGKRPLODPEVOKHATQIIMNMLROEBEELQDILPPCGEISQRTCGRLSVDSQEADSGI 240
 DB 181 TGKRPLODPEVOKHATQIIMNMLROEBEELQDILPPCGEISQRTCGRLSVDSQEADSGI 240

QY 241 DSGTERFPSTSESLAMRNVSLSLPGDLPQTSPIYIARVAQHRRGSDAALLPLNHQGI 300
 DB 241 DSGTERFPSTSESLAMRNVSLSLPGDLPQTSPIYIARVAQHRRGSDAALLPLNHQGI 300

QY 301 DQSPKFLIIGPEEDYDPGYFNNESDIIFODLEKLSHPAYLVYELRYILSQADPGPLFY 360
 DB 301 DQSPKFLIIGPEEDYDPGYFNNESDIIFODLEKLSHPAYLVYELRYILSQADPGPLFY 360

QY 361 LCSEVYQQTNPKDSRLGKDINWIFLEKNAPLRVKIPEMLOAEIDLRLNNEPVRNVLC 420
 DB 361 LCSEVYQQTNPKDSRLGKDINWIFLEKNAPLRVKIPEMLOAEIDLRLNNEPVRNVLC 420

QY 421 AQAVALLEIOEQINDYRSKRTGLGSLYGENDLGLDGPRLRQMAEQALAGIILSK 480
 DB 421 AQAVALLEIOEQINDYRSKRTGLGSLYGENDLGLDGPRLRQMAEQALAGIILSK 480

QY 481 YEEDRSAPMDFAVNTFMSHAGIRLRESRSSCTAEKTQSAPODKKMLPFFPKTKKQSSNSK 540
 DB 481 YEEDRSAPMDFAVNTFMSHAGIRLRESRSSCTAEKTQSAPODKKMLPFFPKTKKQSSNSK 540

QY 541 KEKDALEDKKRNPLIRYIGKPKSSOSIKPGVNRNITIOHFENSHOYDVEPGTQRLSTGS 600
 DB 541 KEKDALEDKKRNPLIRYIGKPKSSOSIKPGVNRNITIOHFENSHOYDVEPGTQRLSTGS 600

QY 601 PPEDLLESOSRSREIRLGRSGSLKGREEMKRSKKAENVPRPSDVMDAAAEARLHQSA 660
 DB 601 PPEDLLESOSRSREIRLGRSGSLKGREEMKRSKKAENVPRPSDVMDAAAEARLHQSA 660

QY 661 SSSASSLSTRLENPTPPPTPKMGRSIESPNLGFCTDYLPHLEDDLGOISDLPEPE 720
 DB 661 SSSASSLSTRLENPTPPPTPKMGRSIESPNLGFCTDYLPHLEDDLGOISDLPEPE 720

QY 721 VQNMQHTVKCKDVANLTQREIDROEYINELFYTEASHLTKRLVDLIFQORAKKEMLMR 780
 DB 721 VQNMQHTVKCKDVANLTQREIDROEYINELFYTEASHLTKRLVDLIFQORAKKEMLMR 780

QY 781 BELARLFPMLPELIEIHNSMCEAMKLRREGPIIRDISPMLARFGPRAEELQOVAOF 840
 DB 781 BELARLFPMLPELIEIHNSMCEAMKLRREGPIIRDISPMLARFGPRAEELQOVAOF 840

QY 841 CSYQVALLELIRTKQKESRFOLFQOEAESHQCRQLQRLDILVSEMRILTKYPLLENI 900
 DB 841 CSYQVALLELIRTKQKESRFOLFQOEAESHQCRQLQRLDILVSEMRILTKYPLLENI 900

QY 901 IKHTBEGTSEHEKLCRAPDQCREILKFEVNAVKQTENRRLRLEGYQKRLDATALERASNP 960
 DB 901 IKHTBEGTSEHEKLCRAPDQCREILKFEVNAVKQTENRRLRLEGYQKRLDATALERASNP 960

QY 961 AAEFSKLDLITTKMHIEGLTWIRISDKRLDQVALLLEDLVVLLOQOEBRLLLKSKSKTA 1020
 DB 961 AAEFSKLDLITTKMHIEGLTWIRISDKRLDQVALLLEDLVVLLOQOEBRLLLKSKSKTA 1020

QY 1021 VSSSDSKQTFSPVLKNAVLINRSVATDKRAFFICTSELGPPQIYELVALTSSDKNIME 1080

DB 1021 VSSSDSKQTFSPVLKNAVLINRSVATDKRAFFICTSELGPPQIYELVALTSSDKNIME 1080

QY 1081 LLEEAQONATKHPGAAPIRIHPSPGQEPAYOGSTSSNVEINDSEVYHTEKEPKLPGC 1140
 DB 1081 LLEEAQONATKHPGAAPIRIHPSPGQEPAYOGSTSSNVEINDSEVYHTEKEPKLPGC 1140

QY 1141 PGPEQRYQDKLIAGEPVOEEDDEELRTLPRAPSPSLDENNGIRTRDVLALGPIILM 1200
 DB 1141 PGPEQRYQDKLIAGEPVOEEDDEELRTLPRAPSPSLDENNGIRTRDVLALGPIILM 1200

QY 1201 EGLADALDVENLRILILMSILPGHTVKTQAAAGEEDLTPPSVSTSHSPMPGSPG 1260
 DB 1201 EGLADALDVENLRILILMSILPGHTVKTQAAAGEEDLTPPSVSTSHSPMPGSPG 1260

QY 1261 QAPTISDSTRLARPEGSGEGEDVAVSSLAHLPPRTSSGVWDSPELDNRPAEAASTEP 1320
 DB 1261 QAPTISDSTRLARPEGSGEGEDVAVSSLAHLPPRTSSGVWDSPELDNRPAEAASTEP 1320

QY 1321 AASYKVRKVSILLPGGVAAKVAAGSNATPDSGQSFSELEVEGGAQATGNCFYVSMGAC 1380
 DB 1321 AASYKVRKVSILLPGGVAAKVAAGSNATPDSGQSFSELEVEGGAQATGNCFYVSMGAC 1380

QY 1381 PLDSTSTPTGTPSPSQCCHSLPAMPTEPQPYRGVRCQCSSLVRRDVYIFHTIQLTIK 1440
 DB 1381 PLDSTSTPTGTPSPSQCCHSLPAMPTEPQPYRGVRCQCSSLVRRDVYIFHTIQLTIK 1440

QY 1441 LHRKDMELAHRELKSLGESSGGTTPVGSFHTTEAARTDYSLPKAEALASQONQ 1500
 DB 1441 LHRKDMELAHRELKSLGESSGGTTPVGSFHTTEAARTDYSLPKAEALASQONQ 1500

QY 1501 EOGSCPEGSDIALDSDATDVASPGP 1527
 DB 1501 EOGSCPEGSDIALDSDATDVASPGP 1527

RESULT 2
 015085 PRELIMINARY; PRT: 1522 AA.

AC 015085;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE KIA0380 PROTEIN.
 GN ARHGEF11/KIA0380.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN:
 RX MEDLINE-9734984; PubMed-9205841;
 RX MEDLINE-99150378; PubMed-10026210;
 RA Fukushima S., Mura C., Zohar M., Iguchi T., Gutkind J.S.:
 "A novel PDZ domain containing guanine nucleotide exchange factor
 RT links heterotrimeric G proteins to Rho.";
 RL J. Biol. Chem. 274:5868-5879(1999).
 CC -i- SIMILARITY: CONTAINS 1 RGS DOMAIN.
 DR EMBL: AB002378; BAA20834.1; -;
 DR HSBP: Q12923; 3PZ;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR000342; RGS.

DR InterPro: IPR000219; RhogEF.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00621; RhogEF; 1.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00233; PH; 1.
 DR SMART: SM00315; RGS; 1.
 DR SMART: SM00325; RhogEF; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50003; PH DOMAIN; 1.
 DR PROSITE: PS50132; RGS; 1.
 SO SEQUENCE 1522 AA; 167703 MW; CA16E125B9F8A4AA CRC64;

Query Match 82.2%; Score 6482.5; DB 4; Length 1522;
 Best Local Similarity 82.3%; Pred. No. 0;
 Matches 1278; Conservative 89; Mismatches 129; Indels 57; Gaps 10;

QY 1 MSIRLPHSIDRSASKKOSHLSPISAWLSLSLSDSTPERTSPSHHROPDSFTSETTAGL 60
 DB 1 MSYLPQSIDR-----LSSLSLSDSAPERRKSPSHHROPDSASETT-GL 43
 QY 61 VORCVITIKOKHGRGFTVSGDRIVLVQSVRPGGAAMKAGKEGDRITIKVGTWNTNSHL 120
 DB 44 VORCVITIKOKHGRGFTVSGDRIVLVQSVRPGGAAMKAGKEGDRITIKVGTWNTNSHL 103
 QY 121 EYVKLIKSGAYAAATLLGSSPSVSGVGLQONPSVAGVLRVNPRIIPPPPPPLPPPOHI 180
 DB 104 EYVKLIKSGAYAAATLLGSSPSVSGVGLQONPSVAGVLRVNPRIIPPPPPPLPPPOHI 163
 QY 181 TGPKPLQDPEVOKHATQILNMLRQEEBELDILPPCGETSORTCEGLRVSQSEADSG 240
 DB 164 TGPKPLQDPEVOKHATQILNMLRQEEBELDILPPCGETSORTCEGLRVSQSEADSG 223
 QY 241 DSGTERPSPISSELMANNVSLSDPGLDSPOTSPYITLARVAQHHRQSDAALPLNQG 300
 DB 224 DSGTERPSPISSELMANNVSLSDPGLDSPOTSPYITLARVAQHHRQSDAALPLNQG 283
 QY 301 DQSPKPLIIPPEEDYDPCYFNNESEDIFQDLKLSHPATLVYLRITLQADPGPLLFY 360
 DB 284 DQSPKPLIIPPEEDYDPCYFNNESEDIFQDLKLSHPATLVYLRITLQADPGPLLFY 343
 QY 361 LCEVYVQOTNPKDSRSLGKDIWIFLEKNAPLRVYKIPEMLQAEIDLRLNNEDEPNVLC 420
 DB 344 LCEVYVQOTNPKDSRSLGKDIWIFLEKNAPLRVYKIPEMLQAEIDLRLNNEDEPNVLC 403
 QY 421 AOEAVMLEIOEJOINDYRSKRTLGSLYGENDLGLDGDPLREBQAEKOLAAGDLSK 480
 DB 404 AOEAVMLEIOEJOINDYRSKRTLGSLYGENDLGLDGDPLREBQAEKOLAAGDLSK 463
 QY 481 YEEDRSAPMDFAVTFEMSHAGIRLRESRSSCTAECTOSAPDKDKWLFPFRTKROSSNSK 540
 DB 464 YEEDRSAPMDFAVTFEMSHAGIRLRESRSSCTAECTOSAPDKDKWLFPFRTKROSSNSK 522
 QY 541 KERDALEDKRNPILRYIKGPKSSSSQ-----IKQGNVNNIIQHEENSHQYDPE 590
 DB 523 KERDALEDKRNPILRYIKGPKSSSSQ-----IKQGNVNNIIQHEENSHQYDPE 582
 QY 591 PGQORLSTGSPFDLLESDSRSRIRLGRSGSLKGREEMKRSKAENVPPRSDVMDAA 650
 DB 583 PGQORLSTGSPFDLLESDSRSRIRLGRSGSLKGREEMKRSKAENVPPRSDVMDAA 642
 QY 651 AEAARLHQSASSASLSTRSLENPPPPPKMGORRSESPNIGCFDVLPHILLEDLG 710
 DB 643 AEAARLHQSASSASLSTRSLENPPPPPKMGORRSESPNIGCFDVLPHILLEDLG 702
 QY 711 QLSDLPEPEVQWQHTVGKDVANLQREIDRQEVINELFVTEASHLRTLRVLDLIFY 770
 DB 703 QLSDLPEPEVQWQHTVGKDVANLQREIDRQEVINELFVTEASHLRTLRVLDLIFY 762
 QY 771 RMKRENAMPREELARLPNLPPELLEIHNKSCAMKRTREGPIITROISDPMLARFGDPAR 830
 DB 763 RMKRENAMPREELARLPNLPPELLEIHNKSCAMKRTREGPIITROISDPMLARFGDPAR 822

QY 831 EELQVAAQCFQSYVALTELIRTKORKESRFQFMQEAESHPOCRRLQDLRLIVSEMQR 890
 DB 823 EELQVAAQCFQSYVALTELIRTKORKESRFQFMQEAESHPOCRRLQDLRLIVSEMQR 882
 QY 891 TKYPLLENIITKTEGGTSEHEKLCRADOCREILKFNVAVKOTENRRHLEGYQKRLA 950
 DB 883 TKYPLLENIITKTEGGTSEHEKLCRADOCREILKFNVAVKOTENRRHLEGYQKRLA 942
 QY 951 TALERASNPLAEFKSLDLTRRMHIEGRLTWMISDKDTLDLVLLEDLVLLOKDER 1010
 DB 943 TALERASNPLAEFKSLDLTRRMHIEGRLTWMISDKDTLDLVLLEDLVLLOKDER 1002
 QY 1011 LLLKCHSKTAVGSSDSKOTSPYVLKNAVLIRSVATDKRAFFITCTSELGPQIYELVAL 1070
 DB 1003 LLLKCHSKTAVGSSDSKOTSPYVLKNAVLIRSVATDKRAFFITCTSELGPQIYELVAL 1062
 QY 1071 TSSDKNIMELLEAVONATKHPGAPIRPHSPGSGEPAVQGSSTSRVEINDSEYHT 1130
 DB 1063 TSSDKNIMELLEAVONATKHPGAPIRPHSPGSGEPAVQGSSTSRVEINDSEYHT 1122
 QY 1131 EKEPKLREGSGPEORVQDKLQIAQGEVQ--EDEBELRTLPRAPSLDGENRGIRTRD 1188
 DB 1123 EKEPKLREGSGPEORVQDKLQIAQGEVQ--EDEBELRTLPRAPSLDGENRGIRTRD 1182
 QY 1189 PVLATLTPGLMEGLADALEVDENLRHLITLMSLLPGHTVKTQOAGPEDDLPPTSPVS 1248
 DB 1183 PVLATLTPGLMEGLADALEVDENLRHLITLMSLLPGHTVKTQOAGPEDDLPPTSPVS 1242
 QY 1249 ITSHPMWPGSGGAP--ISDSTRLARPEGSGEEDVAVSLAHLPPRTSSGVWDSPE 1306
 DB 1243 ITSHPMWPGSGGAP--ISDSTRLARPEGSGEEDVAVSLAHLPPRTSSGVWDSPE 1302
 QY 1307 LDRNPAEAESTERPAASYKVRKVSLLPGGCGAAYAGN--ATPDQSGSELSSEVE 1363
 DB 1303 LDRNPAEAESTERPAASYKVRKVSLLPGGCGAAYAGN--ATPDQSGSELSSEVE 1351
 QY 1364 GGAQATGNCFYVMPAGPLDSTEPGTSPSPQCHSLPAMPTEPOPYRGVRGQGS--- 1420
 DB 1352 GGAQATGNCFYVMPAGPLDSTEPGTSPSPQCHSLPAMPTEPOPYRGVRGQGS--- 1409
 QY 1421 -----SLVRDQVYFHTTEQLTIKHLRLKDMELAHRELKSLGESSGCTTPVGSFHT 1474
 DB 1410 -----SLVRDQVYFHTTEQLTIKHLRLKDMELAHRELKSLGESSGCTTPVGSFHT 1469
 QY 1475 EAARWTDYSIPPAKELASDSONGECSPEGSDIALEDATPTAVSPGP 1527
 DB 1470 EAARWTDYSIPPAKELASDSONGECSPEGSDIALEDATPTAVSPGP 1522

RESULT 3
 O9NZNS PRELIMINARY; PRT; 1544 AA.
 AC O9NZNS; 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE GUANINE NUCLEOTIDE EXCHANGE FACTOR.
 GN Homo sapiens (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PROSTATE;
 RX MEDLINE=20160919; PubMed=10681437;
 RA Kourlas P.J., Strout M.P., Becknell B., Veronese M.L., Croce C.M.,
 RA Theil K.S., Krahe R., Rutu T., Knutilla S., Bloomfield C.D.,
 RA Caligiuri M.A.;
 RT "Identification of a gene at 11q23 encoding a guanine nucleotide
 RT exchange factor: Evidence for its fusion with MLL in acute myeloid
 RT leukemia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2145-2150(2000).


```

QY 580 FENSHOYDVPBGTORLSTGSPFEDLLESDSSSELRIGSGSLKREEMKSRKAENP 639
Db 280 -----PCRRHLK-----VEADE-----KPGADRRKGGSLGMSNRORTG 312
QY 640 RPRSDVMDMAAEEAARLHQSASSASSLSTRLENPTPP-FPPKMRRIESPNIQFCTD 698
Db 313 TFGQD-----NGVSLHLSTDSVDSREP-GVDTPOBGTDPGPGTSLP----- 357
QY 699 VILPILLEDDLCOLS-----DLEPEEYONMOTYVKGDVANLTQREIDR 743
Db 358 -LAPPESTEDNGETESPEFGDDGDEPGRSGLELEPE-EPGMEELVPPDILLSPKQVQR 415
QY 744 QEYNLEFTEASHLTLVLDLIFQMRKREMLPRELARLPPLLEIHNMCW- 802
Db 416 QEYISLTLTEAHVMLVLDLIFQPMADGFFPLDELQNTFSLDELIEVHSLFDR 475
QY 803 AMKKLEEGPIRIDSDPMLARFDGPARBELQOVAQFCYSOVALELIRTKQKESRPO 862
Db 476 LMKRQESYLLIEIGDVLARFDGAGSMFOKISSRFGSROSFALQAKQKRRRC 535
QY 863 LEMQEAESHPOCRRLQRLDILVSEMOQLTKYPLLENIKHTEGTSEHEKLCRADQCR 922
Db 536 AFVQEAESHPRCRRLQKLMITPEMQLTKYPLLOSIGNTEEST-ERKVELAAECOR 594
QY 923 ELTKPNEAVKQENHRELEGYOKRLDATALERASNPAAEKSILDTIRKMHIEGPI 982
Db 595 ELIHNHQAIVRMEDLLRLKDYQRLDLTHLROSSDPMLSEFKNDITKKKVHESPLW 654
QY 983 RISKDKTLDLQVLLLEDVLVLLROBERLLKCHSKTAAVSSDSKQTFSPVLKNAVLR 1042
Db 655 RYTKKDAIEVHLLDDLLLLQRODERLLKSHSRTLPPTDGMKMLRPLRLTSAMR 714
QY 1043 SVATDKRAFFITCTSELGPQIYELVALTSSDKNIMWELLEAVONATKHPAIPRHP 1102
Db 715 EVATDHKAYVIFTWD-QEAOIYELVAQTSERKNCMLITETAGSLKYPADASRLKPR 773
QY 1103 SPPGSGEAPVOSTSRVAINSEVYHTEKEPKLPEG-----QRP- 1148
Db 774 SPSSTREPL-----SSSENGTGAEMAPADARTERLNDLLPCRCPEQLAATALQKVL 830
QY 1149 D-KOLIAGEPVOEDEEELRTLPRAPSLDGEN-RGIRTRDPV-LTALGTLMEGLAD 1205
Db 831 SLKQILLS-----TEEDS-----GAGPRDDGVPGGARPGVHMQEILENLSLEVAI 879
QY 1206 AALDEVE-----NLRHLLMSLLPGHTVKQAAGE 1235
Db 880 RQLEELIEEFCTRLPL--SOLGTLSPMLAPE 911

```

RESULT 6

```

QY 092888 PRELIMINARY: PRT: 912 AA.
AC 092888;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR P115-RHOGEF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN 111
RP SEQUENCE FROM N.A.
RA MEDLINE=96411762; PubMed=8810315;
RA Hart M.J., Sharma S., el Masry N., Qiu R.G., McCabe P., Polakis P.,
RA Bolla G.;
RT "Identification of a novel guanine nucleotide exchange factor for the
RT rho GTPase."
RL J. Biol. Chem. 271:25452-25458(1996).
DR EMBL: U64105; AAB17896.1; -.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RHOGEF.
DR InterPro: IPR003489; Ribosomal_S30.

```

```

DR Pfam: PF00621; RHOGEF; 1.
DR SMART: SM00235; PH; 1.
DR SMART: SM00325; RHOGEF; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
SQ SEQUENCE 912 AA; 102516 MW; 54AE5ACAFC291FIC CRC64;

Query Match 15.2%; Score 1196.5; DB 4; Length 912;
Best Local Similarity 34.2%; Pred. No. 2,3e-66;
Matches 335; Conservative 167; Mismatches 324; Indels 153; Gaps 32;

QY 308 IIGPE-EDYDPCYFNN--ESDIIIFODLEKSHPRYLIVFLYIISODPGLFLYLCSE 364
Db 23 ITGAEDEDFENLENTSEFQNSQFSLQVKKRPAHLQVLAQVLDQEPGLCLCHAD 82
QY 365 VYQQTNPKRSRSLGDIWNIFLEKNAPLVKIPREMOAETD--LRLNNEPRLVLCOA 421
Db 83 MGLSGPKRAKKAFLDFVHSLFKETAVALRVPPVAVAFELDRADLLSEVQRRFVQEV 142
QY 422 QEAVMLEIOEQINDYRSKRTGLGSLYGENDLGID-----GDPLBEROMAEKQIAA 473
Db 143 YSQQVAVGROLEDPRSRKLMGMP--WEQELAQLEAVWGRDASYEARERHVAERLLMH 200
QY 474 LGDI--LSKYEEDRSAPMDPAVNTFEMSHAGIRLSRSRCTAEKTOQAPDKDKMLPFT-- 529
Db 201 LEMQHTTISTDEKSAAYVNAIGLYRHLYGVTKSG-----DKKSGRNFPRK 247
QY 530 -----PKTKROSSKREKDALEDKKNPILRYIGPKSSSOSIKRGVNRITQ 578
Db 248 KVMGRNRSDDPKTKKGLSI--LDAARWNNGEPQVDFRLKAEVAKPG----- 297
QY 579 HFENSHQIDVPEPGTQRLSTGSPFEDLLESDSSSELRIGSGSLKREEMKSRKAENP 638
Db 298 -----AYDRKGVGM-----SRD-----RNIGAPGDTPGVSLHPPL 331
QY 639 RPRSDVMDMAAEEAARLHQSASSASSLSTRLENPTPPFPKMRRIESPNIQFCTD 698
Db 332 DSPREPGADAPL-----GDSFPGPMSLSLAPPESTDGAET-ESPEFG----- 378
QY 699 VILPILLEDDLCQ--LSDLPEEP-EVQNMQHTYVKGDVANLTQREIDROEYINELFTEA 755
Db 379 -----DEGEPGRSGLELEPEPPGMRVLPDTHSLPSKQVQOEYISELVTGA 429
QY 756 SHLRTLRYLDLIFYORMKENIMPREELARLPNLPPELIEHNSKCE-AMKKLEEGPII 814
Db 430 AHVRMLRYLHDLFEQPMACEFFPLEELQNTFPSLDELIEVHSLFDRIMKRROESGLI 489
QY 815 RDISPMLARFDGPARBELQOVAQFCYSOVALELIRTKQKESRFOQFMQEAESHPOC 874
Db 490 EETGVLLARFDGAGSMFOKISSRFGSROSFALQAKQKRRPCAFVQEAESHPRC 549
QY 875 RRLQRLDILVSEMOQLTKYPLLENIKHTEGTSEHEKLCRADQCEILKPYNEAVKQ 934
Db 550 RRLQKQKQIMPEMQLTKYPLLOSIGNTEEST-ERKVELAAECCEILHVNQAVRD 608
QY 935 TENRRLEGYOKRLDATYALERASNPAAEFSLDTTKMHIEGPIYTRISKQDTLDQV 994
Db 609 MEDLLRLDYORRLDLSHLROSSDPMLSEFKNDITKKKVHESPLWRYKDKRAVEYH 668
QY 995 LILEDLVLLVLRQBERLLKCHSKTAAVSSDSKQTFSPVLKNAVLRISVATDKRAPII 1054
Db 669 LILDDLLLLQRODERLLKSHSRTLPPTDGMKMLRPLRLTSAMTEVATDKRAFYVL 728
QY 1055 CTSELGPQIYELVALTSSDKNIMWELLEAVON-ATKHPAARPIHPSPGQEPAYQ 1113
Db 729 FTWD-QEAOIYELVAQVSEKNCMLITETAGSLKYPADASRPKP-RPRSSSTREPLLS 786
QY 1114 GSTSRVAINSEVYHTEKEPKLPEG-----TQPEPRYVD-----KOLIAGEP 1158
Db 787 SSENQN--GGRETSPPARTERTILSDLLPCRCPEQLAATALRKVLSLKQLLF--P 840
QY 1159 VQEEDEEELRTLPRAPSLDGENGIRTRDPVLLATGTLPLMEGLAD-----AALDEVE- 1212

```

Db 841 AEEDN-----GAGPRDGD--GVGGGPLSPARTOE-IOENLLSEETMKELEEE 889
 Oy 1213 ---NLRLILMSILPGHTV 1228
 Db 890 EPCRLRPL--SOLGAGNSV 906

RESULT 7
 ID 092116 PRELIMINARY; PRT; 919 AA.
 AC 092116;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE LSC PROTEIN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Steinbrenner H., Wuensche C., Seisler J.,
 RT "Cloning of a rat homologue of Lsc, a mouse oncogene with structural
 RT similarities to the Dbl family of guanine nucleotide exchange
 RT factors";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ36911; CAA15426.1; -
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR00219; RhGEF.
 DR Pfam: PF00621; RhGEF.
 DR SMART: SM00233; PH; 1.
 DR SMART: SM00325; RhGEF; 1.
 DR PROSITE: PS50003; PH DOMAIN; 1.
 SQ SEQUENCE 919 AA; 102597 MW; CF328CD2587717F8 CRC64;

Query Match 15.0%; Score 1180; DB 11; Length 919;
 Best Local Similarity 33.2%; Pred. No. 2.5e-65;
 Matches 334; Conservative 154; Mismatches 309; Indels 210; Gaps 35;

Oy 307 LITGPE-EDVDPGFYNNESD--IIFODELKSHPATLVFLKRLISQADPGPLFYICS 363
 Db 20 IITGAEDEDENELEAPEDONSOFOLEQRRPAHMLLOVALQFEPGLCLHA 79
 Oy 364 EYVQOTNPKDSRSLSGKIMINIFLEKNAPLRVKIPEMLOAEIDLR--LRNNEPRNVICE 420
 Db 80 DMLSSLPGEKKKAFDLFYHSFLEKTVLKVPPVSAFELDTRPDLISDVORFTQE 139
 Oy 421 AQEAVMLEIOEQINDYRSKRTGLGSLYGENDL---GLD--GDPLRROMAEKQALAL 474
 Db 140 VVGSOQAAYVRQLIEDFRSKRLMGMTPEQELSLLEPWIGDRGVEAREHVRARLLSHL 199
 Oy 475 GDI--LSKYEEDRSAPDFAVNTFMHAGIRLRERSSCTAETQSAPEDKWLPEF--- 529
 Db 200 EEMOHTISTDEEKSAAYVTATISLVRHLGVRTKSG-----DKKSGHPRKK 246
 Oy 530 -----PKTKOSSNSKKEKDALEDKRNPLRYICKPKSSSOSIKPNVNRNIQH 579
 Db 247 VMGRKRDEPPKTKKGS-----IIDPARWNR-----GSPSAPD-----CRH 284
 Oy 580 FENSHQYDVE-PTQRLSTGSPEDLLIESDSSREIRLGRSGSLAKREMKSRKAENV 638
 Db 285 L---KYVEDEKPG-----PAD-----RKGL-----GISSRRRTV 311
 Oy 639 PRPSRDVMDAAAEARLHOSASSASSLSTR--SLENPPP--FTPMGRRSIESPULGF 695
 Db 312 GTPGOD-----NPGVSLH---PLSYDSLDSREGVDTPOPGDTPPGQPTSLP----- 357
 Oy 696 CTDIVLPHLEDDLGL-----SDLEPEP--EVQNMQHTGKGVANVLTQREI 741
 Db 358 ----LAPPESTEDNGETESPEPPGDDGEPGRSGLEQPEDEPPGKRELVPDTLGLPRNOV 413

Oy 742 DROEVINLEFYTEASHLRTLEVLDFIYQRMKKNLMPRELARLPNLPETIEHNSWC 801
 Db 414 KRQEVISELVTETAAHAYMLRVLDLFYQPAEGGFPLEQNIFFSLDELEHSLFL 473
 Oy 802 E-AMKKLREEGPIIRIDISDPMARPDGPARRELOOVAOFCSYOSVALLETIRTKRESR 860
 Db 474 DRLMKRROESGYLIEIGDVLARFDGEGSWFOKISSRCSQSFALQOKAKQREPR 533
 Oy 861 FQLMQAEASHPOCRRLQRLIYSEMOQLTYPLLENIITKTEGGTSEHEKLCRARQ 920
 Db 534 FCAPVOAESRPRCRRLQKMDIPEMORLTKYLLQSIGQNT-EPASRAKVELAAEC 592
 Oy 921 CRELLKPVNEAVKQENRHRLEGYQKRLDAPALERASNPILAEKSLDITTRKMIHGPL 980
 Db 593 CRELLHYNQAVRMEDILRLKDYQKRLDTHLKQSDPMSEKKNIDITKRLVHGSL 652
 Oy 981 TWIRISKDTIDLOVLLLEDDVLLQROBERLLKCHSKTAVGSSSKQTESPVKLNAV 1040
 Db 653 TWRLTKRAVEVHVLDDLDDLLLRQDEGCSSHTSRTILTPPDGTMRLRPVLRISAM 712
 Oy 1041 IRSVATDKRAFFICTSELAPPOIYELVALTSSDKNIMMELLEAVON-ATKHPGAAPIP 1099
 Db 713 TREVATDHKAFFYLTWD-OEAQITELVAQTSERKSCALITETAGSLKVPAPASRPKP 771
 Oy 1100 IHPSPGSEAPAYGSTSSRVEINDESVYTERKEPKKLPGPGPEQGRVQKQILAAGEPV 1159
 Db 772 -RPPSSRTRELLSSSEN-----GTGG-----TEAAPA 798
 Oy 1160 OEDDEBELR-TLPAPPSLDGENRGITRPDVLATLGPLIMEGLADAEVENVRLH 1218
 Db 799 DARTERILINDLPECRPGPEQO-----LAATFALQKVLSTKQIL 836
 Oy 1219 LMSLLPGHTVKTQAAGEPD-DLTPPSVVSITSHMPDPSGQAPR 1264
 Db 837 L-----STEEDSGAGPPRDGVP-----GGAGPEGTHT 865

RESULT 8
 ID 09BSB1 PRELIMINARY; PRT; 853 AA.
 AC 09BSB1;
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE SIMILAR TO RHO GUANINE NUCLEOTIDE EXCHANGE FACTOR (GEF) 1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MUSCLE, RHABDOMYOSARCOMA;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC005155; AA05155.1; -
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR00219; RhGEF.
 DR InterPro: IPR003489; Ribosomal_S30.
 DR Pfam: PF00621; RhGEF; 1.
 DR SMART: SM00233; PH; 1.
 DR SMART: SM00325; RhGEF; 1.
 DR PROSITE: PS50003; PH DOMAIN; 1.
 SQ SEQUENCE 853 AA; 95941 MW; 9426FC3E4CFD706C CRC64;

Query Match 14.6%; Score 1151; DB 4; Length 853;
 Best Local Similarity 34.2%; Pred. No. 1.5e-63;
 Matches 320; Conservative 166; Mismatches 313; Indels 136; Gaps 32;

Oy 342 VVELRITLSADGGLFYICSEYVQOTNPKDSRSLSGKIDIMINIFLEKNAPLRVKITBMDQ 401
 Db 1 MALLQHVALLQFEPGLCLHADMLSLGPKKAKKAFDLFYHSFLEKTVLKVPPVPPNVA 60


```

Db 649 RLTSAMTREVATDHKAFYVLTWD-QEAOIYELVAQIVSERKNNMSALITETAGSLKVPAP 707
QY 1094 GAAPILPHSPSGQEPAYOGSTSSRVEINDSEVYHTEKEPKLPEGPGEORVQOKOLI 1153
Db 708 ASRRKP-RPSPSSREPLSSSEN----- 730
QY 1154 AOGPEVOEDEELRT-----LPRAPSLDGENRGIRTRDPVLLATGPLMEGLADA 1206
Db 731 --GNGGETSPADKRTIRILSDLLPFCRPGREGO-----LAAT 766
QY 1207 ALDEDVENRLHLLMSLLPGHTVKTQAGEPED-DLPTPSVSTSHPMWPGSPGQAPTI 1265
Db 767 ALKRVLSLKOL-----LEPAE--EDNGAGPRPDGDPGPGG-----PRSPARTQEI 810
QY 1266 SDRRLARPEGSGQEPGEVAVSSLAHLPRTRSSGVWDSPDL 1307
Db 811 QEN--LRLLETMTKLEEL-EEFCRLRLPILLSQGLTSLPSL 849

```

```

RESULT 10
096F17
ID 096F17 PRELIMINARY; PRT; 879 AA.
AC 096F17;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:19628).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011726; AAL11726.1;
SQ SEQUENCE 879 AA; 98767 MW; 1D0863A5D1A57C9B CRC64;

```

```

Query Match 14.48; Score 1139; DB 4; Length 879;
Best Local Similarity 33.5%; Pred. No. 8.6e-63;
Matches 326; Conservative 165; Mismatches 309; Indels 172; Gaps 35;

QY 308 IIGGE-BDVPDGYFNN--ESDITFODELKLKSHRAYLVYLRITLSQADGPGILFYLCSE 364
Db 23 IIGAEDEDENLEETNSEEQNSQFSLQVKKRPAHLMALLOHVALQFEQPLV----- 76
QY 365 VYQQTNPDKDSLSKIDWNIIFLEKNAPLRVYKIPMLQAEID--LRLNNEDPRNVLCSEA 421
Db 77 -----LRYVPVPPNVAFFELDRTRADLISDYGVRFFQEV 109
QY 422 QEAVMLEIQIQINDYRSKRTLGSLGYENDLGLD-----GDLPEROMAEOLOA 473
Db 110 VOSQQVAVGQQLDEFRSKRLMGMP--WEQELQLEAMVGRDRASVBARRHVAERLLMH 167
QY 474 LGDI--LSKREEDRSAMDFAVNTFMSHAGRLRESRSSCTAEQTSAPRKDKMLPRFPK 531
Db 168 LEEQHTITSTDEKSAVVAIVAGLYMRHLVRTKSG-----DKKSGRNF-- 212
QY 532 TKROSSSKKEKDALEDKKNPILRYI---GKPKSSSOSIKPQNVNIITOHFEN-SHOY 586
Db 213 RKKYMGNRSDDEPAKTKKGLSILDAARMNGEPQ-----VPDFRHLKAEV 256
QY 587 DVEPEG-TQLSTGSPEDLLESDSSSEIRLGRSGSLKREKRSKRAKNVPRPSDV 645
Db 259 DAEPGATDRKGGVGMF-----SRD-----RNIGAPQDTPGVSLHPLSDPDREP 305
QY 646 DMAAAARLHOSASSASSLSTRLENPTPTPKMGSRSTESPLGCTDVIPLHL 705
Db 306 GADAPLEL-----GDSSPQGMPSLESILAPPESTDEGAET--ESPEPG----- 345

```

```

QY 706 EDDLGO--LSDLEPER-EVONMOHTVQKDVYANVLTQREIDROEYINELFYTEASHLTTR 762
Db 346 --DEGERGRSGLEPERPEPPGMRELVPPDILHSPKQVQARQEVISLLVTEAHYRMLR 403
QY 763 VLDLIFYQRNKENLMREELARLPNLPPELLIEHNSWCE-AMKLTREEGPIIRIDISDP 821
Db 404 VLHDLFPQMAECLEFPLEELQINFPSLDELIEVHSLFLRLMKRRQESGLYLEIGDVL 463
QY 822 LARFDGARELDQVAAOFCSYSALELRTQKRSRQLTMOQKESHPQCRRLQDR 881
Db 464 LARPDAGESWFOKTSIRSFCSQSFALQKAFORQDPFCALVQEAESRPRKRLQDK 523
QY 882 LIVSEMORLTKYPLLLENIITKTEGCTSEHEKLCRARDQCREILKFVNEAVKOTENHRL 941
Db 524 MIPTEMORLTKYPLLLOSIGQNTDEPT-EREKVELAECCREILHHNVQAVRMEDLRL 582
QY 942 EGYOKRLDATALERASPLAERKSLDTLRKMIHSGPLTWIRSKDITDLOVLLLEDV 1001
Db 583 KDYORRLDLSHLROSSDPMLESEFNKNDITKKKLVHSGPLTWIRVYTKDAVEVHLLDDL 642
QY 1002 VILOROEELLKCHSKTAAGSSDSKOTFSPLKINAVLIRSVATOKRAFFICTSELGP 1061
Db 643 LILQRODERLLKSHSRTLPPTPDGKTMRLPVLRLSAMTREVATDHKAFYVLTWD-QE 701
QY 1062 POLYELVALTSDKNIMWELLEAVON-ATKHGAAPIPHPSPSGQEPAYOGSTSSRY 1120
Db 702 AQIYELVAQIVSERKNNCALITETAGSLKVPASRRKP-RPSPSSREPLSSSEN- 759
QY 1121 EINDSEVYHTEKEPKLPEG-----PGEPORVOD-----KQIIAAGEPVQEDDEE 1165
Db 760 --GGERETSPADARTERILSDLLPFCRPGREGOLATRLKRVLSLKOLLF--PAEDBN- 812
QY 1166 ELRLTPRAPSLDGENRGIRTRDPVLLATGPLMEGLAD-----AALEPVE-----NLRH 1216
Db 813 -----GAGPRPGD--GVGGGFLSPARTQE-IQENLLSLEETMKOLEEELSEFCRLR 863
QY 1217 LILMSLLPGHTV 1228
Db 864 LL--SQLGNSV 873

```

```

RESULT 11
ID 015086 PRELIMINARY; PRT; 750 AA.
AC 015086;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, last sequence update)
DE KIAA0382 PROTEIN (FRAGMENT).
GN KIAA0382.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=9734984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002380; BAA20836.1; -.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhGEF.
DR Pfam; PF006621; RhGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhGEF; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
FT NON_TER 1
SQ SEQUENCE 750 AA; 84746 MW; 8B14AE449C26CE72 CRC64;

```


Query Match 13.0%; Score 1022; DB 4; Length 750;
 Best Local Similarity 32.8%; Pred. No. 1.4e-55;
 Matches 269; Conservative 132; Mismatches 232; Indels 188; Gaps 25;

QY 750 LEVTEASHLTLVLDLIFYQMRKKNLMPREELARLPNLPELIEIHNSWCEAMKLE 809
 DB 1 LFYTERAHVHTLKVLDQVDFQVRVRSREGIILSPSELARKIFSLIEDIILQHLGLNQMKAVR 60
 QY 810 --EGPIIRDISDMLARFPGPAREELOOVAOCYSQVSALEDFIRKORRESFOLFMOE 867
 DB 61 RNFSTYIDQIGEDLWFSFGPEEKLKHAATFCSNQPFALEMTKSRÖKDSRFÖFVD 120
 QY 868 AESHPÖCRRLQRLDLIVSEMQRLTKYPLLENIKHTEGTSEHEKLCRARDOCREILKF 927
 DB 121 AESNPICRRLOKDIIPQOMRLTKYPLLENIKHTEGTSEHEKLCRARDOCREILKF 179
 QY 928 VNEAVKOTENRHRLEGVOKRDLATALERASNPAAEFKSIDLTRKMIHEGPIITWISND 987
 DB 180 VNOAVEAEKKORLEDYÖRLDITSSLKSEYVNEVELRNLDTKRMHIEGPIVWVKVND 239
 QY 988 KTIIDLVLLEDVLVLRQEOERLLKCHSKTAVGSSDSKÖPSPVLTKNALVIRSVATD 1047
 DB 240 KTIIDLVLLEDVLVLRQEOERLLKCHSKTAVGSSDSKÖPSPVLTKNALVIRSVATD 299
 QY 1048 KRAFFITCTSELGPQIYELVALTSSDKNIMWELAEVQONATKHPGAAPRIHPSPG- 1106
 DB 300 NKALFVISMDSNG-AQIYELVAQTVSEKTYWÖDLICR--MAASVEQSTKPIPLPOSTPGE 357
 QY 1107 -----SQEPAYÖSTSSRVEIND----- 1124
 DB 358 GDNDEEDPSKLEEOHGISVTGLÖSPDRDLGELSTLISSKPOSHSTSGKSEVRLFYA 417
 QY 1125 ---SEVTEKEPKKLPEGS-----PGEPÖR-VÖD-----KÖLLAÖGEVÖE 1161
 DB 418 ERQFAEQHTDGLKEVGDYÖAIIPDSHLPVSEERWALDALRNGLKÖLLVQÖGLTE 477
 QY 1162 ED-EELKRLPRAPSLDGENNGICIRTRDYLALITGPILMIEGLADVEDVNLRLILM 1220
 DB 478 KSVÖEWMÖHPR-----YRT-----ASÖGP-----QTDSEVIONSENK--AY 512
 QY 1221 SLIPGHTVQAAEGEDDLTPPSVVSITSHMPDPSGQAPITISDTRL----- 1271
 DB 513 HSEGMPTGTGDIATCYSPKTSFSFA--PRD--SYGLAPÖDSQASNIILVMDHMTF 568
 QY 1272 ----ARPEGS--ÖPEGEDVAVSSLAHLPPR--TRSSGVWDSPELDNR----- 1310
 DB 569 PEMPTMEPEGGLDSEHFPDAREAHSDENPSGDAVKNKEKDVNLRTISGNVLLDGYD 628
 QY 1311 PAEAASTEPASIKYKRVVSVLLPGGCVGAAYAGSNA--IPDSGÖSESL-----E 1361
 DB 629 PÖESSTDEAVS-----SLTÖPMTGIPAVESTHÖQHSPONTHSDGAIPTPEFLVÖ 683
 QY 1362 VEGGAÖTNCFCYVWAPRLDSTREPTGTPPSQCHSLPAMPTEPPQYRGVGGQCSS 1421
 DB 684 ÖRMGAHEY--SCREIÖSPSSCADSQÖ----- 708
 QY 1422 LVRRDVIYFHTLEÖTLTKRLKÖMELARRELLKSIGES 1462
 DB 709 -----IMEXYTHKEADLEHLKKEVESYTIICÖRLAASA 741

RESULT 12
 096BF4 PRELIMINARY; PRT; 540 AA.
 AC 096BF4
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 61.3 KDA PROTEIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PRIMARY B-CELLS FROM TONSILS;
 RA Strausberg R.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC015652; AAH15652.1;
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 540 AA; 61278 MW; BD25E9E5662DE009 CRC64;

Query Match 12.2%; Score 962.5; DB 4; Length 540;
 Best Local Similarity 42.0%; Pred. No. 4.6e-52;
 Matches 227; Conservative 105; Mismatches 160; Indels 49; Gaps 16;

QY 714 DLEPEVQNMÖHTGKÖVAVNLQÖREIDÖRVINELPVTESHRLTLYVLDLIFYÖMR 773
 DB 17 ELEPE-EPPGMRELVPDDTLHSLPXSÖVKÖEYISLELVAHAHVMRLVHLDTFFÖPWA 75
 QY 774 KENLMPREELARLPNLPELIEIHNSWCE--AMKKLREEGPIIRDISDMLARFDPAREE 832
 DB 76 EÖLFPLEÖLÖNIFPSLDLEIVHSFLDRMKRQESYILEELGDVLARFDGAEQSW 135
 QY 833 LÖQVAÖFCSYÖSALELIRTKÖKRESFÖLFMÖAESHPÖCRRLQRLDLIVSEMQRLTK 892
 DB 136 FÖKISRSFCSRÖSFALÖULKAKÖRKDPRCFCAVÖEAESRPRCRRLOKMIPTEMÖRLTK 195
 QY 893 YPLLENIKHTEGTSEHEKLCRARDOCREILKFVNEAVKOTENRHRHIEGVÖKRLDZTA 952
 DB 196 YPLLÖSIGÖNTEEP--EREKVELAÖCCÖRETLHNÖVNRÖMEDLRLKÖDÖYÖRLDUSH 254
 QY 953 LERASNPAAEFKSIDLTRKMIHEGPIITWISKDTLÖVLLEDVLVLRQEOERLL 1012
 DB 255 LÖSSDPMSERKNDITKRLKVLHIEGPIITWISKDTLÖVLLEDVLVLRQEOERLL 314
 QY 1013 LCHSKTAVGSSDSKÖTSPVLTKNALVIRSVATKRAFFITCTSELGPQIYELVALTS 1072
 DB 315 LKSHSTLITPTPDGKTMRLPVLRTSAMTREVATÖHKAÖYVFTWD--ÖEAQIYELVAQTV 373
 QY 1073 SDKNIMWELAEVQON-ATKHPGAAPRIHPSPSQÖEPAYÖSTSSRVEINDSEVYTE 1131
 DB 374 SERKMWALITFASGLKVPAPASRKP--RPSSTRELLSSSNGN---GGRITSAD 429
 QY 1132 KEPKKLPEGS-----PGÖRÖVÖD-----KÖLLAÖGEVÖEDEBELTLPAPPS 1176
 DB 430 ARTERILSDILPFCRPPGEGQLAATALRKVLKÖLLF---PAERDN-----GAGPP 478
 QY 1177 LDGNGRGTTRDYPVLLATITGPILMIEGLAD-----AALÖVE-----NLRLITMSLPGHT 1227
 DB 479 RDGD--GVPGGGLSPARTÖE--IÖENLISLETMKÖLEELPEFCRLRPL--SÖLGGNS 533
 QY 1228 V 1228
 DB 534 V 534

RESULT 13
 044381 PRELIMINARY; PRT; 2559 AA.
 AC 044381
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SHAR PEI/DRHÖGEF2.
 GN RHÖGEF2 OR CG9635.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=9808790; PubMed=9428514;
 RA Barrett K., Leptin M., Seltman J.;
 RT "The Rho GTPase and a putative RhoGEF mediate a signaling pathway for
 the cell shape changes in Drosophila gastrulation."
 RL Cell 91:905-915(1997).
 DR EMBL; AF032870; AAC38820.1; -.
 DR HSSP; P29476; 10AV.
 DR FLYbase; FBgn0023172; RhoGEF2.
 DR InterPro: IPR001525; C5_DNA_mech.
 DR InterPro: IPR002219; DAG_PE-bind.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001849; PH.
 DR pfam; PF00130; DAG_PE-bind; 1.
 DR pfam; PF00595; PDZ; 1.
 DR pfam; PF00169; PH; 1.
 DR pfam; PF00621; RhoGEF; 1.
 DR SMART; SM00109; C1; 1.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; RhoGEF; 1.
 DR PROSITE; PS00094; C5_MTASE_1; UNKNOWN_1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KM Phorbol-ester binding.
 SQ SEQUENCE 2559 AA; 280857 MW; 112CCEFF3362BCDF CRC64;

Query Match 12.0%; Score 944.5; DB 5; Length 2559;
 Best Local Similarity 20.9%; Pred. No. 6.9e-50;
 Matches 482; Conservative 279; Mismatches 651; Indels 895; Gaps 82;

QY 17 QSLHSPIASWSSLSLSDSTPERTSPSHRPSDSTETAGLVQRC-----VII 67
 DB 204 QSNPSNVLQAPGERSLNLTPLSRDLSGHTQESTPTATPTPSLALPKNFQYLTLTV 263
 QY 68 OKDOHFGFTVSGDRIYLVQSVRPGAGAMKAGVKEGDRITKNGTWTNSHLEVYKLIK 127
 DB 264 RDSNRTGKAKVSGDNVFEVSVKPGCAAEIAGLVAGDMILRVNGHEVRLKHPYVGLIK 323
 QY 128 SGAYALTLTGSS---PSPVGV-----SGLQONPSYAGVLRV----- 161
 DB 324 ASTTVLAVKRSQKLTSPSSSVYTPSTPILSGRDTASTITGPVDSIKRREMETYKIQ 383
 QY 162 -----NP----- 163
 DB 384 TLQKMLEQKLNLERLKSQDNPSYKLSSEANIRKLEQLHQVGAEDAPYVKLQAAAGNKN 443
 QY 164 --TIIP-----PP-----PP----- 170
 DB 444 TALLTPNOIOLHSATSHSQGFHLLHHNNHNNNTPPQOQASTSPALSLILPSLSS 503
 QY 171 ----- 170
 DB 504 LSLGTRKNTEKDLTSSPGLTDFLQOQRMGSHQAESMSQSMHQTSTPTSOQFFPHQ 563
 QY 171 -----PPPLP---PPQ-----HITG 182
 DB 564 QOHREKETGPTSGKKNKFLISRLIEDVPPPLPORNPPQLNLDLKNAGSFGSHLVA 623
 QY 183 P-----KPLDPEYKATTOILNM 202
 DB 624 PVSDDLRATSPOLNRSGOQOOLPSTDNSPSNAKSKSKITKALSDP---KMTQMLQX 680
 QY 203 ----- 202
 DB 681 ESASAGAAGSIEVDGPPPLPPLPGMTEDMSRGSCNLAQPNVSTAFNPLVSTT 740
 QY 203 -----LROEEELDQIL-----PPGGEFSQR----- 223
 DB 741 TAVQNDNLNIAFPLSORPNIVQOLQYQOQOQHOMSGGATGALGOTPNLGNKKNHRRVGS 800

QY 224 -----TCEGRLSVDSQEADGLDSTGERPSESIM-----NRNS 259
 DB 801 SPDNMHRHPRDRTTKTSSGWEIYENDGESS-PPGTPPLPYLSSIMHTVLEDDENNRGA 859
 QY 260 VLSDPG--LDSPQ-----TSVILARVAQH--HRROGSDALLPLNHQIDOSPPLIT 309
 DB 860 AAGPGVFIESHOFTPWAGASSPILSHSHMAAOSND-----TOKEIIS 906
 QY 310 GREPDVD-----PGYFNESDIIFODEKLKSHPAYLVFLRYILSOADGPPILFYL 361
 DB 907 MEDNSDLDEPFTDENGPEFNLLRLAE-----NVTFAIFLNVYISNDPAPLFYL 960
 QY 362 CSEVYOOTNEPKDSRLGKIDWNIIFLEKNAPLR-VKIPEMLOAIDLRLRNEDP-----RN 416
 DB 961 ITELYKEGSKDMRKNAEYEHSTFLVPRLASWYRQDESLAREVDVNLQLEYKVELKRT 1020
 QY 417 VLCEAOEAVMLEIQEOINDYRSKRTIGLSLYGEND--ILGLDGPPLEROMAEKOL--- 471
 DB 1021 VFLRSRRRAKDLISEQLREFQOKRTAGLGYGPTTDKLAETDKLRE-QIIDKYLMPN 1079
 QY 472 -----AALGDIISKYEEDSAPDFA--VNTFMSH----- 499
 DB 1080 LHALIEDENGSPEDYRKVALCALSTVIYRIENTRPPSSIVERVHHFVSROKSEKSRIT 1139
 QY 500 -----AGIR-----LRE-----SRSCCTAEKTS 518
 DB 1140 MGKNRKNNVAGHPLVLRQYEVYHCHNCQRTIINGVSPQGYHCTDKNLTIRQCSKYVDES 1199
 QY 519 APDKKWLFPFPKTKKOSNSKKEK-----DALEDKRR----- 551
 DB 1200 CPGP-----LPOAKRLAHNDKISKFMGKIRPRTSDVIGNEKRSRODELDVELTPDRGQ 1253
 QY 552 -----NPLIRYIGKPKSSQSIKGNVNNIIQHENSQYVPPPG----- 592
 DB 1254 ASTVROPDDRPPANISIRKNGTSCNTSGL---NTTDLOSSFHGSCANDSINPGGAGC 1310
 QY 593 TORLST-----GSFPE-----DLESSESR-----SEIRLGRS 620
 DB 1311 NMDLSTSVASTPTSTGSGVANAAGLSAFELNALDTPVDEKARERYSQHPKSAFVSNRS 1370
 QY 621 GSLGKREEMKRSKRAEN-----VPRPSVDVMDAAAEARLHQASASSASSLS----- 668
 DB 1371 ESKYKERLSNKRNNRSRKRTSDPSLSRP-NDEQDLGLSNATYVGSNSLSAGSGESP 1429
 QY 669 TRSLNPTRP-----FTPMGRSSTESPML-----GCTDYILPHLEDDLGQSS- 713
 DB 1430 STSMENHPAAGAGVQVPPMGLNQNHPLHLIQHAAQOYCQODSFQAGLAGAAGSSAAS 1489
 QY 714 -----DLEPEPEVO--NMQHTVCGDVANLTQREIDREIYINELFVT 753
 DB 1490 NSSFWNAGHPLPARWTLSEDEDDVNEADMSMWVAEVALALTDAEKKQOETINELYOT 1549
 QY 754 EASHLRTLRVLDLIFYORMKENLMPREELARLP-PNLPELIEIHNSCEAMKLR-EEG 811
 DB 1550 ERNHVRLTKLIDLFLPLYESGLISQDHLFLPPALSLREIHGAFESQILKORLEHN 1609
 QY 812 PIIRDISDPMLARFGDARELQOVAOCSYQVALIELTRKOKESRROLMOEAESH 871
 DB 1610 HVAVNTIGDLADMFGQGVVLCFAAQCARQOIALEAKERKNDEMLOKLKLSSESH 1669
 QY 872 POCRRLOLRDLIVSEMOQLTKYPLLENITKHT-----EGGSTSEHEKICARADOCREIKT 927
 DB 1670 KACRRLEKRLPLTYQRLTKYPLLENITKTYVRLPENTTEAEVIAQRAVESKRLIVE 1729
 QY 928 VNEAVKOTENRRHRLLEGYQKLDATALERASNPPLAAEFKSIDLTRKMIHEGPLTRISKD 987
 DB 1730 VNOAVKTAEDAKHLQNIQRLDRSSYDK-----EEKKKIDLTQHNHIIHGNGT--IKKN 1781
 QY 988 KTLIDQVLLLEDLVLLQROEEFRLLK-CHSKTAVGSSDSKQTFSPVLTKNA-VLIRSA 1045
 DB 1782 PSVOLGELLFENMIVLLTKODKYYLKNLHTPLUST-----TNKVPSPYMSIDAOTLIRQEA 1837

```

OY 1046 TDKRAFFICTSELGPQIYELVALTSSDKNIMWELLEEAVONATKH----- 1092
DB 1838 ADRKNSFFLI---KMTSOMLELRAPSSSECKTWFHFSDVAAROSKNRKNASNNHDTSI 1894
OY 1093 ---PGAAPPIHSPPGSOD----- 1118
DB 1895 SGPALALAP-HSNTTESLESLSTDYOPPLAATLTTPPLAPMLPATYTPADPATNNSNVS 1953
OY 1119 -----RVEINDSEVYHTEKEPKKLEGGPEORVODKOLI-AOGEPOV-----EED 1163
DB 1954 SLTGVLQRPORDATASESDADYVNTPKPRSSONEVNRFTMSIRSGEPLOKYSANGTEAN 2013
OY 1164 EEBELR-----TLRPRAPSLDGENR-----GIRTRDPVLLAL-----TGLMEGLA 1204
DB 2014 DVTALRHSSOSTRESVRSGSTGEERNSTYGVANGSKRDSASYCSNNSNNTRLTLMOISPLV 2073
OY 1205 D-ALEDEVENLRHLLMLSLPGHFTVKTQAGEPED-----DLTPRP-----SVASI 1249
DB 2074 DFTALQVSTISPAHNTAPVLTPEBEKLRRLDASTIRNDLEKOKITICDIFRLPVEHYQIYDI 2133
OY 1250 TSHPMWPGSPGQAPRTISDSTRLARPEGSG-----PEGEDVAVSSL---AH- 1291
DB 2134 AMMP-----EAPKDSADIALAAYDQIOTLTKMLNEYMHVPEQEVSAVSTAVGCHHE 2186
OY 1292 -----LPRTRSSGVWDSPELDNRNPAEAASSTEPASIVVVRKVSLLPGCGGGAAYV 1343
DB 2187 KEKLRKKVAPSSSFSSPPPLPPRHOHAQAQIIPP---RLMKRLQTLDDV----- 2238
OY 1344 AGSNAIPDSGSESELSVEVEGGAQATGNCFYVMPAGPLDSSTPTPTSPSCCHSLPA 1403
DB 2239 ---AIHEDDOCYCEIDELR-----LPAIPSKPHERRP-PLAP----- 2272
OY 1404 WTEPOPIRGVGGGQSSSLVRDNDVIHPTT-----EQLTLKLRL- 1444
DB 2273 FNTPEKTSQSV---IDASKRQSTDAVPEGLLEQEPLEGGDKETKTEGDENEVTVPSSDKLS 2328
OY 1445 ---KDMELAHNELKSLGSESSGGTTPVGSFHTEARMTDLSLPPAKELALASONSQO 1500
DB 2329 ESCNEROCVEADITREKVAADPTTSKNEAASVDELPSOSREIKTAKENASKSVADKEDNE 2388
OY 1501 EOGSCPEEGSDIALEDSATDTRAVSPGP 1527
DB 2389 E---TIEKGV-ASTVDSSTQTSPTESP 2411

RESULT 14
OY 09V7X3 PRELIMINARY: PRT: 2559 AA.
AC 09V7X3:
DT 01-MAY-2000 (TREMBLrel. 13. Created)
DT 01-MAY-2000 (TREMBLrel. 13. Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19. Last annotation update)
DE RHOGEF2 PROTEIN.
GN RHOGEF2 OR CG9635.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephyryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyryotida; Dirosophillidae; Drosophilla.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```

```

RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Duthin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris W.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nuskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003804; AAF57915.1; -.
DR HSSP: P29476; 10AV.
DR FLYBASE: FBgn0023172; RhogEF2.
DR InterPro: IPR001525; CS_DNA_meth.
DR InterPro: IPR002219; DAG_PE_bind.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001849; PH.
DR InterPro: IPR00219; RhogEF.
DR Pfam: PF00130; DAG_PE_bind; 1.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhogEF; 1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhogEF; 1.
DR PROSITE: PS00094; CS_MTASE_1; UNKNOWN_1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50106; PDZ; 1.
SQ SEQUENCE 2559 AA; 280805 MW; 07725DE04982605C CRC64;

```

```

Query Match 11.9%; Score 936.5; DB 5; Length 2559;
Best Local Similarity 20.8%; Pred. No. 2.2e+49;
Matches 481; Conservative 279; Mismatches 652; Indels 895; Gaps 82;

OY 17 OSHLSPLASMLSSLSGDSPTPERTSPSHRQPSDSTETAGLVORC-----VII 67
DB 204 QSDNSNPVLOAPGERSLINTPLPSHLSGHTQESTPATTPSPSLAPKPNFOYLTLT 263
OY 68 OKDOHGFETVSGDRIVVQSVRPGAAKAGVKEGDRIRKNGVTMNSSHLEVVKLIK 127
DB 264 RKDSNGVMKYSGDNPNVVESEYKPGGAELIAGLVAGDMILRVNGHEVLEKHPYVGLIK 323
OY 128 SCAYAAALTLGLSS---PPSVGV-----SGLOQNPSSVAGVLRV----- 161
DB 324 ASTVYELAVKRSOKLRTSSSVSVYTPSPILSGRDRFASITGPPQVDSIKRREMYTIQ 383
OY 162 -----NP----- 163
DB 384 TLQKMLEQEKINLERLKSQDNPNPSYKISEANIRKLREQLHQVGAEDAPTVKLAAGNKN 443
OY 164 --IIPP-----PP----- 170

```


DE PUTATIVE GUANINE NUCLEOTIDE EXCHANGE FACTOR RHOGEF2.
 GN RHOGEF2 OR CG9635.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_TaxId=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Haecker U., Perrimon N.;
 RL Genes Dev. 0:0-0(1998).
 DR EMBL: AF031930: AAB8816.1; -
 DR HSSP: P29476; 1QAV. -
 DR Flybase: F89N023172; RHOGEF2.
 DR InterPro: IPR001525; C5_DNA_mech.
 DR InterPro: IPR002219; DAG_PE-bind.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001649; PH.
 DR InterPro: IPR000219; Rhogef.
 DR Pfam: PF00130; DAG_PE-bind; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00169; PH; 1.
 DR Pfam: PF00621; Rhogef; 1.
 DR SMART: SM00109; C1; 1.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00233; PH; 1.
 DR SMART: SM00325; Rhogef; 1.
 DR PROSITE: PS00094; C5_MTASE_1; UNKNOWN_1.
 DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
 DR PROSITE: PS50106; PDZ; 1.
 KW Photol-ester binding.
 SQ SEQUENCE 2559 AA; 280946 MW; 7BC661AA1E6E840E CRC64;

Query Match 11.7%; Score 922.5; DB 5; Length 2559;
 Best Local Similarity 20.8%; Pred. No. 1.7e-48;
 Matches 479; Conservative 279; Mismatches 654; Indels 895; Gaps 82;

QY 17 QSHLSPIASWLSLSLSDSTPERTSPSHRQPSDSTETAGLVQRC-----VII 67
 DB 204 QSDNSNPVLQAGERSLNLTLPSRLSGHSGHQSSTTPATPTSTPLALPKNFQYLTLV 263
 QY 68 QNDQHFQFTVSGDRIVLYQSVRPGAAKAGVKBEDRIKNGMTWYSSHLVEYVKKLK 127
 DB 264 RQDSNGYGMKVGSDNPVEYEVKPGGAELAGLVAGDMILRVNGHEVRLKHPYVGLIK 323
 QY 128 SGAYVALTLTGSS---PSPGV- -SGLDQNPVAGVLRV----- 161
 DB 324 ASTVEYELAVKRQOKLTRPSSVSVYTPSTPLTSGRDKRTASTGPPQVDSIKRREMYKIQ 383
 QY 162 -----NP----- 163
 DB 384 TLQKMLEQKLEMLERKSDQNNPSYKLSSEANIRKLEQLHQYGAEDAPTIVKLOAAAGNKN 443
 QY 164 --TIIP-----PP----- 170
 DB 444 TALLTPNOIOLASATHSNOQFHLHHNNLHNNNYPPQOQOAPASTPAFLSLRSLSS 503
 QY 171 -----PP----- 170
 DB 504 LSLGTRKKTKEDLTSSPFGLTDFLOQQRMSHOAESMSQSMHQHTTPTSOQFFPHQ 563
 QY 171 -----PPLP--PPQ-----HITG 182
 DB 564 QOHREKFGPTSGKNNKFLISRSLIEDVPRPLPRQNRPLRLDLKNGNAPGSHLVA 623
 QY 183 P-----KPLDDPEVQKATQILMM 202
 DB 624 PVSDDLRAATPOLNRSOQOOLPRSTDNSPSNAKSKRSKIKTKALSDP---KMSDQMFLOM 680
 QY 203 ----- 202

DB 681 ESASAGAAGSIEVDGPPPLPRPLPGMMTEDMRSQCNLAQNPSTAFNPLVSTT 740
 QY 203 -----LRQEEELQDIL-----PPCGEFSOR---- 223
 DB 741 TAVQNDNLNIAFPPLSORPNIYQQLQYQQQQQHOMSGGATGALGOTPHLGKNNKHVGS 800
 QY 224 -----TCBGRLSVDSQADSGLDGSTERPSSISELM-----NRNS 259
 DB 801 SPDNMHRPRDRITTKTSSWEIWEKDESS--PPGTPPPYSSSHMTVLEDPNENNRGA 859
 QY 260 VLSDDP--LDSPQ-----TSPTILAVAOH--HRQSGDALPLLNHGIDQSPKPLTI 309
 DB 860 AAAGPCVFIESHQFTPMACASSPIPSLSNMHAAQSD-----TQKEIIS 906
 QY 310 GPEEDYD-----PGYPNNSDIIFQDLKSHPAVLVFLRYILSQADPGPLFLYL 361
 DB 907 MEDENSDDLDEPFIDENGPPNNLTRLEAE-----NVTFALAFVNISSDPAPLFLYL 960
 QY 362 CSEVIQOTNPKDSRSLGKIIMNIFLEKNAPLR-VKIPEMLQAEIDLRLANNEDP-----RN 416
 DB 961 ITELKESGSKDMRKWAVEIHSTFLVPAPRLSMYRQDESLAREVNVLDLEYDKVEILRT 1020
 QY 417 VLCEQEAVMLEIQIINDYRSKRTLGSLXGEND--LLGIDGPILREROMAEKQL--- 471
 DB 1021 VELRSKRRAKDLISEQLREFOQKRTAGLSTIYGPTDQKLAETDKLRE-QIIDKYLAPN 1079
 QY 472 -----AALGDILSKYEEDRSAPMDFA--VNTFMH----- 499
 DB 1080 LHALIEDENGSPPEQRYKVALCSALSTVYRIENRPPSSLYERNHHVSDKSPKSR 1139
 QY 500 -----AGIR--LRE-----SRSCITAEKTOS 518
 DB 1140 MGNKRKNVNGHPVLRYQYEYTHGCHNCOTIIGVSPQGYHCTDCKLNIRQSKVYDES 1199
 QY 519 APDKQMLPEFPPTKKOSSNSKREK-----DALEDKR----- 551
 DB 1200 CGCP-----LQAKRLANDKISKFMKIRPTSDVIGNERSKQDELEINELTPDQ 1253
 QY 552 -----NPLRYIGKPKSSOSIKPGVNRNIIQHFENSHOYDVEPG----- 592
 DB 1254 ASIVQPSDRRDPANISIRSNQNTSCNTSGL--NTTDLQSSFHGSCANDSINPQAGAC 1310
 QY 593 TORLST-----GSFPE--DLESOSR-----SEIRLGRS 620
 DB 1311 NMDLSTVAASTPTSTGSAAGLSAPAEINADLYDVKERRRRYQHPKHSAPVSMKS 1370
 QY 621 GSLGREEMKRSRKAEEN-----VPRPSDYDMAAEEAARLIHQSSASSSLS----- 668
 DB 1371 ESKYERLSKRRNRNRKRTSDPSLSRP--NDEQDLGLSNATYVGSNSLSAGCTESP 1429
 QY 669 TRSLENPTPP-----FTPKMGRSIESPNL-----GECTDYILPHLLEDLQOLS-- 713
 DB 1430 STSMEHFAAPGAAGVQVPMPGLNQNHPHLIQHQAQYCCQODSFOGAGAAASSAS 1489
 QY 714 -----DLEPEVQ--NMQHTYQKDVANLQREIDROEVINEVLT 753
 DB 1490 NSSFNAGHPPLVAPRWLTLESEDEDVDVNEADMSMYAAEVSALTLTAERKROEITINEIYT 1549
 QY 754 EASHRLTAVLDLIFYORARKENLMPREBLARL--PNLPELEIINSNCEAKKRLR--BEG 811
 DB 1550 ERNNHYRTKLKDLRFLPLRYESGLSODHLLLFPPALLSLREIGHARPGQSIKORRIENH 1609
 QY 812 PIIRDISPMIARFDPAPARELQYVAAQFCYSQVSALELIRTKOKRESRFQLEQAESEH 871
 DB 1610 HVTNITIGDLADMFDGSGGVVLCFPAQFCAPQOIALALEKRRKDMLOKLKKSSEH 1669
 QY 872 PQCRLOLRDLIVSMOULTKYRILLENIKHT-----BGGTSEHKRLORADQCEILKF 927
 DB 1670 KACRLELKDLPLVLORTKYRPLLENDYKTVARLBLENTEAARIDAPVAESSRILVE 1729
 QY 928 VNEAVKOTENRRLLEGYQKRDLATALERASNPDLAEFSLDTTRKMHHEGPIYTRISK 987
 DB 1730 VNGAVRTAEDAKHLQNIQRKDRSSYDK-----BEFKKLDLTQNHLLHDGNLT--IKKN 1781

QY 988 KTLDLOVLLLEDVLLVROBERLLK-CHSKTAVGSSDSKOTFSPVYLKINA-VLIRSVA 1045
Db 1782 PSVOLHGLEPMMIVLLTKODDKYLLKNLHTPLSI-----TNKVPSPIMSIDADTLIRQEA 1837
QY 1046 TDKRAFFIICTSELGPPQIYELVALTSSDKNINMELLEAVONATKH----- 1092
Db 1838 ADKNSFPLI---KKKTSQMLERAPSSSECKTFKHFSDVAAROSKRNKRSKNASNHDTSI 1894
QY 1093 --PGAAPRPHSPPGSQE-----PAYQSTSS 1118
Db 1895 SDPALAIIP-HSNTKESLELSTDTVOPLAATATLTTPLAPMLPIATVTPAPATNNSNVS 1953
QY 1119 -----RVEINDESEVYHTEKEPKLPEGPPEQRYODKOLI-AQGEVVO-----EED 1163
Db 1954 SLTGVOLRNQORADATASESDADYVNTPKPRSSQNEVNRKMSIRSTGEPIQKYSANGTEAN 2013
QY 1164 EEBELR---TLPRAPSLDGENR-----GIRTRDPVLLAL-----TGPLIMEG-LA 1204
Db 2014 DVTLRHSQSTRESYRPGSTGEERNSTYGMVGNKSRDSASIVCSNNSNNTRTLIMQSPLY 2073
QY 1205 D-AALEDEVENLRHLLMSLPGHTVKTQAAGEPED-----DLPTFP-----SVYSI 1249
Db 2074 DPTAIQVSIPTAHTAEVYLPGEKLRDLASIRNDLLEKOKIICDIFRLPEHYTDQIVDI 2133
QY 1250 TSHPMWPGSPGOAPTISDSTRLARPEGSQ-----PEGEDVAVSSL---AH- 1291
Db 2134 AMMP-----EAPKDSADIALAAYDQIQLTKMLNEYMHVTPQGEVSAVSTAVCGHCH 2186
QY 1292 -----LPPTRSSGVWDSPELDNRPAAEAASSTEPASIKYVRYKVSLLPGGCVGAAKV 1343
Db 2187 KEKLRKKVAPSSSPSPPLPPNROHQAQAQIIPPS---RLMPKLOTLDLDEV----- 2238
QY 1344 AGSNATIPDSQSESELSVEVGGAAQATGNCFYVSMAGPLDSTPEGTTPSPSQCHSLPA 1403
Db 2239 ----AIHEDDDGICEIDELR-----LPALPSKPERPT-TPLAP----- 2272
QY 1404 WPTPEOPYRGVGGQCSLVRRDVIYFHTI-----EQLTIKLRHL- 1444
Db 2273 FNTPEPKTSQSY---IDASKRQSTDAVPEGLLEQEPLEGDKTETKGEDNEVKTVPSDXLS 2328
QY 1445 ----KQMEHLAHRLLKSLGGESSGCTTPVGSFHTEAARMTDYSLSPPAKELASDSQNGQ 1500
Db 2329 ESCNEERQCEVADITKEVADPTTSKNEAASVDELPSQSREIKTAENASKSVADKKEONE 2388
QY 1501 EOGSCPEEGSDIALEDSATDTAVSPGP 1527
Db 2389 E---TIEBGV-ASTVDSSTOTSPTESP 2411

Search completed: September 18, 2002, 10:39:25
Job time: 207 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 10:37:59 ; Search time 16.83 seconds
(without alignments)
3513.057 Million cell updates/sec

Title: US-09-695-795-4
Perfect score: 7883
Sequence: 1 MSIRLPHSIDRSASKKQSHL.....EGSDIALDSATDPAVSPGP 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	521	6.6	1693	1	RIP2_MOUSE
2	509.5	6.5	893	1	GEPL_HUMAN
3	462	5.9	596	1	LFC_MOUSE
4	428	5.4	424	1	LBC_HUMAN
5	318	4.0	1805	1	NEST_RAT
6	262.5	3.3	1356	1	ROM2_YEAST
7	253.5	3.2	2842	1	APC_RAT
8	247.5	3.1	1618	1	NEST_HUMAN
9	246.5	3.1	847	1	VAV3_HUMAN
10	245.5	3.1	843	1	VAV_RAT
11	244	3.1	847	1	VAV3_MOUSE
12	240	3.0	845	1	VAV_MOUSE
13	238.5	3.0	3038	1	TRIO_HUMAN
14	237	3.0	738	1	ECT2_MOUSE
15	233.5	3.0	845	1	VAV_HUMAN
16	232	2.9	868	1	VAV2_MOUSE
17	231	2.9	883	1	ECT2_HUMAN
18	231	2.9	960	1	FGD1_MOUSE
19	229	2.9	2845	1	APC_MOUSE
20	227	2.9	878	1	VAV2_HUMAN
21	227	2.9	2774	1	MAPA_RAT
22	226.5	2.9	2805	1	MAPA_HUMAN
23	226	2.9	1591	1	TIAM_MOUSE
24	223.5	2.8	1155	1	ROM1_YEAST
25	223	2.8	2230	1	GCG4_HUMAN
26	221.5	2.8	519	1	ARHS_HUMAN
27	219	2.8	2044	1	SIF2_DROME
28	219	2.8	2064	1	SIFL_DROME
29	214.5	2.7	1591	1	TIAM_HUMAN
30	213.5	2.7	961	1	FGD1_HUMAN
31	211.5	2.7	753	1	YAF9_SCHPO
32	211.5	2.7	1427	1	REST_HUMAN
33	211	2.7	1781	1	AKAC_HUMAN

34	209.5	2.7	1174	1	ZO2_CANFA
35	209.5	2.7	2738	1	PGCV_RAT
36	206	2.6	1190	1	ZO2_HUMAN
37	204.5	2.6	854	1	CC24_YEAST
38	204	2.6	2464	1	MAPA_MOUSE
39	201.5	2.6	3321	1	KEND_HUMAN
40	201	2.5	4687	1	PLEL_RAT
41	199.5	2.5	2349	1	TPR_HUMAN
42	199	2.5	1505	1	CURT_HUMAN
43	197	2.5	2017	1	MSN_DROME
44	197	2.5	4473	1	PLEL_CRIGR
45	195	2.5	1395	1	SP41_YEAST

ALIGNMENTS

```

RESULT 1
ID RIP2_MOUSE STANDARD: PRT: 1693 AA.
AC P97433:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Rho-interacting protein 2 (Rho-guanine nucleotide exchange factor)
DE (RhoGEF) (RIP2).
GN RHOP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97344280; PubMed=9199174;
RA Housa B., Moolenaar W.H.;
RA Gebbink M.F.B.G., Kranenburg O., Poland M., van Horck F.P.G.,
RT "Identification of a novel, putative Rho-specific GDP/GTP exchange
RT factor and a Rho-binding protein: control of neuronal morphology.";
RL J. Cell Biol. 137:1603-1613(1997).
CC - FUNCTION: GUANINE NUCLEOTIDE EXCHANGE FACTOR THAT INTERACTS WITH
CC RHOA, BUT NOT WITH RAC OR CDC42. ACTIVATES RHOA TO PROMOTE
CC CYTOSKELETAL CONTRACTION AND INHIBIT NEURITE OUTGROWTH.
CC - TISSUE SPECIFICITY: HIGHLY ENRICHED IN THE BRAIN.
CC - SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
CC - SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC - SIMILARITY: TO MAMMALIAN GUANINE NUCLEOTIDE REGULATORY FACTOR LFC
CC AND HUMAN LBC PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U73199; AAB18197.1; -
DR MGD: MGI:1349443; Rho1p2.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR Pfam: PF00130; DAG_PE-bind; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR SMART: SM00109; CI; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhoGEF; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.

```

DR PROSITE: P550010; DH_2; 1.
 DR PROSITE: P550003; PH_DOMAIN; 1.
 KW Guanine-nucleotide releasing factor; Coiled coil.
 FT DOMAIN 99 266 LEU-RICH.
 FT DOMAIN 266 310 COILED COIL (POTENTIAL).
 FT DOMAIN 279 310 POLY-GLU.
 FT DOMAIN 315 318 PHORBOL-ESTER AND DAG BINDING.
 FT DOMAIN 632 698 PH.
 FT DOMAIN 846 1041 COILED COIL (POTENTIAL).
 FT DOMAIN 1095 1184 PH.
 FT DOMAIN 1421 1522 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1693 AA; 190325 MW; 4E96087C449FF14C CRC64;

Query Match 6.6%; Score 521; DB 1; Length 1693;
 Best Local Similarity 20.2%; Pred. No. 9.7e-17;

Matches 290; Conservative 218; Mismatches 496; Indels 430; Gaps 55;

QY 205 QEEELQDILPFGERTSQTCERLSVDSQADSGDSTERTPPTISESLMKNVLSDP 264
 DB 283 EDEAKTEKATPSG-----AAETEEVYRNLESG--RSPSEEE--DAKSIKQ- 326
 QY 265 GLDSPQTSFVLARVAQHRRQGSADALLPLNHQIDQS--PKPLIIGPEEDY----DP 317
 DB 327 -VDGP-----SEH---EDQDLALDRSPDGLKSKHVPASLAAGQLSVLNGDB 372
 QY 318 GFVN-----NESDIIEDLEKLSHPAYLVFLRYILSQADPGFLFYLCSEVYQQTN 370
 DB 373 VVANCWVIDQVGLDINDYINLEGLSTH-----TS 401
 QY 371 PMDSRLGKDINWIFLEKNAPLKVKIPMLQAEIDLRLNNEPRN-----VICGAE 423
 DB 402 PESGRM-----LGPOACHHTLPPTSPGGRPL--IENSETLDAASQSQSVTPSSSR 452
 QY 424 AVMLEIQ-----EQINDYRSKRTILGLSLYGENDLLGLDGPPL--RRROMAKOLAL 474
 DB 453 TSLNLSFGLHPEKEQSHLKKRSSSLDALVADSEGGSEPPICVAVGSSQSPRTGLPS 512
 QY 475 GDLKSYEEDRSAPMDFAVNTFS-----H-----AGIRLRKRS--SCTAEKTQAPDK 522
 DB 513 GDLDSFEFTNPEDCNISHTESLSLSTLHKSLSLIGIR--SRYSQSSPKRISGSR 569
 QY 523 -----DKWLPFPKTKKSSNSK----- 540
 DB 570 LVADFTVCSTSEORYSQOEPPGEKRIQEEENDEVIRAKSSEKXKYSRTSFILMNM 629
 QY 541 -----KEKDALEDKRRN-----PIIYRIGPKS--SSQSITPGNVNIIQH 579
 DB 630 TSPRNKSKMKNKDKTEKEMNRHQFVPGTFSGVLCQSGCDKTLGKESLQACMK- 684
 QY 580 FENSHQ--YDVERPGTORLS-----TGSFPEDLLESDSRSREIRLGRS-----GSLKG 625
 DB 685 -ANTHKGCKDAVPCTCKKQEKYKNNKPPQSLIGSSSVRVPPAPGLSLHPSSMPLGAPG 743
 QY 626 REE-----MKRS--RAEENVPFRSDVMDMAAEARLHQSA--SSSASSISTRSLE 673
 DB 744 RKEFAQVHPLSSVGTLESFRRAVTSLESGDSWRBSHSDLEFGMSGSPSTESF- 802
 QY 674 NPTPTPRKMGRRSIEPMLGCTDYLPHLLEDDIGQLSDLEPEPEVONMOTYGVKDVV 733
 DB 803 -----MMEDVVDSSLMID-----LSSDAQEEAEASLSVIVDPSPFC 837
 QY 734 ANLTQREIDROEVINLFTVASHLTLVLDLIFQYRRKELKMLPRELARLFRPLPL 793
 DB 838 SHOEKVIRQDVIETELMOTEVNHIOTLLIMSEVFRKGMKEELQDLHSTVDKIFPLDEL 897
 QY 794 IETINSMCAKMLREGE-----IIRDISPMPLARDPGARELEQVAAQFCYSQSV 846
 DB 898 LETHRFFFSMKERQESGAGSRNFVINQIGILVQOQSEEMAKMKRIIGFCFHNE 957
 QY 847 ALELITRKQKRSRQLEQWAEASHPOCRRLQIRLIVSEMRQILKYPLLENITKHTEG 906
 DB 958 AMSLFLKELQONK-KFQNFIRKIRNSNLLARRRGIPCCILVTQGITKYPYLVERILQYKE 1016

QY 907 GTSEHEKLCRADQCEIILKFVNEAVKQTEENRRHLEGYCKRLDATALERASNPAAEFKS 966
 DB 1017 RTEEHRDLCKALGLIKDMAADLVKSEYKQKWEILNKENTYTKLKNGHVFRKA 1076
 QY 967 LDUTTRKMIHEGPIWIRISKDTLDQVLLLDLDDVLLQROEBRLCLKHSTAVGSSOS 1026
 DB 1077 LLSQERALLHDGLVTKVITGRKRDLALLDLVLFLEOKDKYTF-----AAVQ 1128
 QY 1027 KOTFSFVLKLNANVLIRSVATDRAPFIICTSELGPQIYELVALTSSDKNIMELLEAV 1086
 DB 1129 KPS---VISLQGLIAREVANEEKGMFLIASSAG--PEWEIHTNKSKEENMMRRIQAV 1184
 QY 1087 QNATKHGGAAPLPIHSPPGSQBPAYQGSTSRKVEINDSEVYHTEKEPKKLDGEGPEER 1146
 DB 1185 ESCPEEEG-----GRTSE-----SDEERRKAEARVAKI-----QQ 1214
 QY 1147 VODKOLIAQGEVQDEDEELRTLPRAPSLDGENGRTROPVLLALGLPLMEGLADA 1206
 DB 1215 COE-IISNODQICYLEKHL-----IYAEIGEL-----S 1244
 QY 1207 ALEDVENLRHLILMSLLPGHTVKTQAGPEDDLPTPSVVSITSHPMDPGSGAPTI- 1265
 DB 1245 GFEDVHLERHLLI-----KP-----DPGEPQAAASLL 1271
 QY 1266 -----SDSTRLA-----RPEGSS-----QPEGEDVAVSSLAHLPPR--TRSS 1299
 DB 1272 AALRAEASLQVAVKASKMDVYSSQSESPGGTVMIDPTSDVPAASPASLVTEGTEGR 1331
 QY 1300 GWDSPSELDRNPAEAASSTEPASVYKVRKVSLLPGGVGAARVAGSNAIPDSGQ----- 1354
 DB 1332 GCMGV-----DGLQGVIVDLAVS-----DAGEVEYR 1359
 QY 1355 ---SESELSEVEGAQATGNCIFYVSMAPGLDSS-----TEPTGTPPSQCHSLPA 1403
 DB 1360 SFGSSQSEIIQAIQNLITRLVLSLQAALTIQDSHIEHKLVLQORESLAPS---HSFRG 1415
 QY 1404 WPEEQPYRGVAGGQCSSLVRRDVIDIEFTIEOLITIKLRKLMELAHRELKS 1457
 DB 1416 GPLDDEKSRVLEKOREELA--NIHKLOHOFQOEBORMRHRTCDQOOREOAOES 1467

RESULT 2
 GEFH_HUMAN
 AC Q92974; Q15079; 075142; PRT; 893 AA.
 DT 30-MAY-2000 (Rel. 39; Created)
 DT 30-MAY-2000 (Rel. 39; Last sequence update)
 DT 01-MAR-2002 (Rel. 41; Last annotation update)
 DE GEF-H1 protein (proliferating cell nuclear antigen p40).
 GN LFP40 OR KIAA0651.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99074271; PubMed=9857026;
 RA Ren Y., Li R., Zheng Y., Busch H.,
 RT "Cloning and characterization of GEF-H1, a microtubule-associated
 RT guanine nucleotide exchange factor for Rac and Rho GTPases.";
 RL J. Biol. Chem. 273:34954-34960(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RC MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 RN [3]
 RP REVISIONS.

RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.;
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 630-893 FROM N.A.
 RX MEDLINE-89168219; PubMed-246560;
 RA Reddy A.B., Chatterjee A., Rothblum L.I., Black A., Busch H.;
 RT Isolation and characterization of complementary DNA to proliferating
 RT cell nucleolar antigen P40.";
 RL Cancer Res. 49:1763-1767(1989)
 CC - FUNCTION: STIMULATES GUANINE NUCLEOTIDE EXCHANGE OF RAC AND RHO
 CC BUT IS INACTIVE TOWARD CDC42, TC10, OR RAS. BINDS TO RAC AND RHO
 CC PROTEINS IN BOTH THE GDP AND GUANOSINE 5'-3-O-(THIO)TRIPHOSPHATE-
 CC BOUND STATES WITHOUT DETECTABLE AFFINITY FOR CDC42 OR RAS. MAY
 CC HAVE A DIRECT ROLE IN ACTIVATION OF RAC AND/OR RHO AND IN BRINGING
 CC THE ACTIVATED GTPASE TO SPECIFIC TARGET SITES SUCH AS
 CC MICROTUBULES.
 CC - SUBCELLULAR LOCATION: COLOCALIZES WITH MICROTUBULES THROUGH THE
 CC CARBOXYL-TERMINAL COILED-COIL DOMAIN.
 CC - SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
 CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC - SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAIN.
 CC - SIMILARITY: COULD BE THE ORTHOLOG OF MOUSE LYMPHOID BLAST CRISIS-
 CC LIKE 1 (LFC ONCOGENE).
 CC - CAUTION: THE SEQUENCE SHOWN HERE COMES FROM THE FIGURE OF REF.1,
 CC APPARENTLY THE SUBMITTED SEQUENCE IS WRONG.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL: U72206; AAC97383.1; ALT_SEQ.
 DR EMBL: AB014551; BAA31626.2; -.
 DR EMBL: X15610; CAA33634.1; -.
 DR InterPro: IPR002219; DAG_PE-bind.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR000219; RhGEF.
 DR Pfam: PF00130; DAG_PE-bind; 1.
 DR Pfam: PF00169; PH; 1.
 DR Pfam: PF00621; RhGEF; 1.
 DR SMART: SM00109; CI; 1.
 DR SMART: SM00233; PH; 1.
 DR SMART: SM00323; RhGEF; 1.
 DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE: PS00081; DAG_PE_BIND_DOM_2; 1.
 DR PROSITE: PS00741; DH_1; FALSE_NEG.
 DR PROSITE: PS50010; DH_2; 1.
 DR PROSITE: PS50003; PH_DOMAIN; 1.
 KW Guanine-nucleotide releasing factor; Phorbol-ester binding; Zinc;
 KW Coiled coil.
 FT DOMAIN 40 86 PHORBOL-ESTER AND DAG BINDING.
 FT DOMAIN 234 431 DH.
 FT DOMAIN 485 571 PH.
 FT DOMAIN 589 610
 FT DOMAIN 797 868 COILED COIL (POTENTIAL).
 FT CONFLICT 1 21 MSRLSLTRARIDRSRLASK -> LGSAAAGCCCCGCCP
 FT CONFLICT 193 193 E -> EA (IN REF. 2).
 FT CONFLICT 579 579 S -> P (IN REF. 2).
 FT CONFLICT 867 867 Q -> P (IN REF. 2).
 FT CONFLICT 885 893 SSPQAMPCT -> SLPGAGALYLSFNPQPSRGTDRLLDPV
 FT (IN REF. 2).
 SQ SEQUENCE 893 AA: 101173 MW: 80AC8FA7F762E9C3 CRC64:

Query Match 6.5%; Score 509.5; DB 1; Length 893;
 Best Local Similarity 23.5%; Pred. No. 1,4e-16;
 Matches 193; Conservative 148; Mismatches 316; Indels 165; Gaps 29;

OY 594 ORLTGSPFEDLLESDSSSRSEIRLGRSGSLKREEM--RSRAEYVPRPSVDMDAA 651
 Db 118 ERPSAIVP-----SDSFROSL-----LGSRRGSSLSLAKSVSTYIAHFND-----E 162
 OY 652 EAARLHQASSSASSLSTRSLENPPPTPKMGRRSIESPNIGCFCDVILPHLEDDLQ 711
 Db 163 SPGLRRILISQSDISMNR-----NRT-----LSVES-----LIDEEVISEL-- 200
 OY 712 LSLPEPEP-----VQNMQHYGKVVANLQRELDROEVINEFVTASHLRTLRVLDLF 768
 Db 201 MSDFENDEKDFADSWSLVNDSSFLQOHKKEVKKQODVYELQTLHHVRLTKMTRLF 260
 OY 769 YQMRKENMPRELRPELNPPELLEINNSWCEAKKIREEG--P-----IIRDISP 820
 Db 261 RIGMLEELHLEPVGVLPCVDELDIRFLSQLERRRQALCGSTRNRYTHLGLD 320
 OY 821 MLARFDGAREELQVAAQFCYSQVSALELITRKORKESEFPQLEMAESHPOCRRLQAR 880
 Db 321 LLSQFGSPSAEQCKTYSFECSSRHSKALKYKELVARDKRFQOFIRKVTIRPAVLKHHGY 380
 OY 881 DLVSEMRUTKYPLLLENIKTEBGTSEHEKLCAROCRRILTFVNEAVKOTENRR 940
 Db 381 ECTLVTRITKYPILTSRLQSHGIEERODLTALGLVKELSNVDEGIYQLKGAR 440
 OY 941 LEGYQRRLDATALERASNPAAE--FKSLDTTRKMIHESGPLRWRSKDKTLDIQLVLE 998
 Db 441 LGITVNRMD-----PRAQTPPGKPGFGRELLARKLIHGOCLMKATNGFKVLLWMT 496
 OY 999 DLVYLQROEBELLKCHSTAVGSSDKQTSFVYKLNANVLRSAVTDKRAFFITCTSE 1058
 Db 497 DVLVFLQEKDQKYPFLDKPS-----VVSLONLIVRDJANOEKGFLLISAA- 543
 OY 1059 LGPQIYELVALTSSKNITMELLEAVONATKHPGAAPPIHPSPGSOEPAYOGSTSS 1118
 Db 544 -PPEMYEVHTASRDRTSTWIRYIQOSVRCPSREPSLIE-----TEDEVYLRIMK 594
 OY 1119 RVEINDSEVYHTEKEPKKLPEGGPQRYODKQLIAGEPVQ--EEDDEELRLTPRAPPSL 1177
 Db 595 ELQQRKRALVELLRE-----KVLGFAEMTHFQAEEDGGSGMALPTLP--- 636
 OY 1178 DGENRGIRTRDPVLLALTEPLMEGLADALEVENMRLHILMSLPLGHTVKTQAAGEPE 1237
 Db 637 ---RGL-FRSESLSPRG-----ERLLQDAIREVEIKLDLY-----GCG 672
 OY 1238 DDLTPPSVSTISHPWDC--SP-----GQAPTISDSTRLARPEGSGPBGEDVANVSSLA 1290
 Db 673 VELLLTPREPALPLEPDGGNTSPVTANGAEATFNGSIELCRADS----- 718
 OY 1291 HLPPTRSSGVWDSPELDRLPAEAASSTEPASVYKRVKSLDPG--GGVGAAKVAGSNA 1348
 Db 719 -----DSSQDRNR-GNQLRSPQELAQRLVNLGLGHGIAQAAVAAQDITLMEAR 765
 OY 1349 IPDSGQSESELSEFV---EGGAQATGNCIFYVSMAPGLDSSTE 1387
 Db 766 FPEGPERREKLCRANGRDGEAGRAG-----AAVAPAEKQATE 802
 RESULT 3
 LFC_MOUSE STANDARD; PRT; 596 AA.
 AC 060875; 009115;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Lymphoid Blast crisis-like 1 (LBC'S first cousin) (Oncogene LFC)
 DE (RHOBIN).
 GN LBCL1 OR LFC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBTaxid=10090;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE=Myeloid;
RX MEDLINE=95355462; PubMed=7629163;
RA Whitehead I., Kirk H., Tognon C., Trigo-Gonzalez G., Kay R.;
RT "Expression cloning of lfc, a novel oncogene with structural
RT similarities to guanine nucleotide exchange factors and to the
RT regulatory region of protein kinase C.";
RL J. Biol. Chem. 270:18388-18395(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Olofsson B.;
RL Submitted (JAN1997) to the EMBL/Genbank/DBSI databases.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS, WITH THE EXCEPTION OF LIVER
CC TISSUE. LEVELS ARE HIGH IN HEMOPOIETIC TISSUES (THYMOS, SPLEEN,
CC BONE MARROW) AS WELL AS IN KIDNEY AND LUNG.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -1- SIMILARITY: TO HUMAN NUCLEOTIDE EXCHANGE PROTEIN LBC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U28495; AAC5234.1; -
DR EMBL: X95761; CA65067.1; -
DR MGI:103264; Lbc11.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhGEF.
DR Pfam: PF00130; DAG_PE-BIND.1.
DR Pfam: PF00169; PH_1.
DR Pfam: PF00621; RhGEF_1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhGEF; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM.1; 1.
DR PROSITE: PS00081; DAG_PE_BIND_DOM.2; 1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene;
KW Phorbol-ester binding; Zinc.
FT DOMAIN 40 86 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 236 433 PH.
FT DOMAIN 487 573 PH.
FT DOMAIN 493 496 POLY-LEU.
FT CONFLICT 1 21 MSRIESTLRARIDRSKEQATK -> MSGNRQPSRRKG
FT CONFLICT 1 21 (IN REF. 2).
FT CONFLICT 156 156 A->V (IN REF. 2).
FT CONFLICT 574 596 MISSING (IN REF. 1).
FT SEQUENCE 596 AA; 68564 MW; E168508BDC7C6E13 CRC64;
SQ

```

Query Match 5.9%; Score 462; DB 1; Length 596;
 Best Local Similarity 26.2%; Pred. No. 1.3e-14;
 Matches 119; Conservative 113; Mismatches 173; Indels 50; Gaps 11;

```

DB 257 IMTRFRTGMELEQMEPEVVOGLPCVDELSDIHTRFLNQLERRRQALCGSTRNFVI 316
QY 815 RDISPMILAFDGPAREELQOVAOCYSQVALLEIRFKOKESRPFOLFMOEASHPOC 874
DB 317 HRLGDLISQFSGSNAEDQKRTYSEPCSHHTKALKIKLYARDKRFQFIKMTRSAYL 376
QY 875 RRLQRLDLIVSEMRQITKYPPLLENIKHTEGTSEHEKLRAPDOCREILFVNEAYQ 934
DB 377 KRHGVOECILLVQRTTKYKPVILINRLQNSHGVEEYQDLASALGLVKELLSNVQDVHE 436
QY 935 TENRRHREGYOKRLDATALERASNPAAE--FKSIDITRRKMHGPIWTRISKNTIDL 992
DB 437 LKEARLQETVRRMD---PRAQTPVPGKPGKRDLELRKILHRCGLIMKTATGFKDV 492
QY 993 QVLLLEDDLVLLQROBERLLKCHSKTAVGSSDSKQTESPVKLNAVLRVATDKRAEF 1052
DB 493 LLLMTDVLVFLQERDQKIFSLDKPS-----VYSIQLNLIROIADNAQKMF 540
QY 1053 IICTSELGPPQIYELVALTSSDKNIMELLEAVQ 1087
DB 541 LISS---GPPEMYEVMAASRDDRTWIRVIOQSVR 572

RESULT 4
LBC_HUMAN STANDARD; PRT; 424 AA.
AC 012802;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE LBC oncogene (P47) (Lymphoid blast crisis oncogene).
GN LBC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94119604; PubMed=8290273;
RA TOKSOZ D., Williams D.A.;
RT "Novel human oncogene lbc detected by transfection with distinct
RT homology regions to signal transduction products.";
RL Oncogene 9:621-628(1994).
CC -1- FUNCTION: STIMULATES EXCHANGE ACTIVITY ON RHO PROTEINS IN VITRO,
CC BUT NOT ON CDC42, RAS OR RAC. MAY BIND CALCIUM IONS.
CC -1- TISSUE SPECIFICITY: RESTRICTED TO HEMATOPOIETIC CELLS AND SKELETAL
CC MUSCLE, LUNG AND HEART. NOT FOUND IN BRAIN, PLACENTA, LIVER,
CC PANCREAS OR KIDNEY.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -1- SIMILARITY: TO MAMMALIAN LFC ONCOGENE AND TO MOUSE RHO-GUANINE
CC NUCLEOTIDE EXCHANGE FACTOR (RGEF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U03634; AAC50065.1; -
DR MIM: 605396; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhGEF.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhGEF; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhGEF; 1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.

```



```

OY 766 LIFYOMRKENIMPREELARLPNPELIEIHNMSCEAMKKLEEGPIIRDISDPMALAR 825
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 818 V-----EEEDGRIVKPLEKVSQ-----DSLGSLAEENV-----845
OY 826 DGPARELOQVAOFCYSVALERIKOR-----ESR---FQJFMOAESHPCCRL 877
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 846 -----OPLRYLEEDDCINS-----LLEDKTHKSLGLEDRNGDSIIIPDESETOVSLRPP 896
OY 878 QLRDLIVSEMOURLTKYPLLENIKTEGTSHEKLCARADCCREILKFNVAEKOTEN 937
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 897 EEDD-----QRTVINLEKESQEFSSSEED-----VMSRSLEGEN 933
OY 938 RHRLEGYOR--RIDATALERASNPLAEFKSLDITRKMIHESPLTWISKDKTDLQVL 995
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 934 HSLISSEVEREOMVESQLEKESQDSG---KSLEDESEQETFF--GPL-----EKEVAESLRSL 984
OY 996 LLEDVLVLLQROERLLKCHSKTAVGSSDSKOTFSPVLKINAVILRSVATKRAFFIC 1055
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 985 AGOD-----OEOKLEDETOOTLRAVGEOMAVSPREKVDLEKPLGNODE-----I 1032
OY 1056 TSELGPPQIYELVALTSSDKNIV-----MELLEAVONATKHPGAPI-----1098
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1033 ANSLGKENQESIVSL--KEKGIEYKSLFETIEIPELFAEEDLERKKSIDTOEPLMSTEV 1090
OY 1099 -----PIRPSPPGSEPRAYOGSTSRVELINDSEVYHTEKEKKLPE--GP--GPEORVQOK 1150
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1091 ARETYEPPEDEPPGS-----LGS---VDBNRETLTSLKEQSELSLKWVNETVEEDS 1141
OY 1151 QLIAGEP--VOEDEDEELRTLPRAPPSLDGENR-----GIRTRDP 1189
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1142 QOCLQYEBEGLQROESIREYKQELPSSGNQORWEDYVKGAVGQEARLATTVGTVEIK 1201
OY 1190 VLLATLGPLMEGLADALEDEVENIRHLI--LMSLLPGHTVTKQAAGEPDDLTTP-----1242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1202 AELHLRG-----OGGEERAAAEGLLODIVGEAMSL-----GSSEPKQGVPAALD 1248
OY 1243 -----TPSVSTISHMWDPSPGQAPITISDTL--ARPESQEGEDVAVSSLA 1290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1249 NLEGALVPAQSMPEVEREDRAQAQEDQSIETVLGLEAARGLELEQVAGLEDRP 1308
OY 1291 H-----LPPRTSSGWMSPELDNRPAEAASSTEPASVYKVRKSLLPGGVGAKYA 1344
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1309 HFAREARIP-----SLGE--ESYAKAKIAQGLEPKER-----KEGALDSGLIELPKIS 1357
OY 1345 GSNALIPDSQSESESEVSEGAQATGNCFYVMPAGPLDSTEPRTGTPSPSQ-----1397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1358 -SEALECQCHHESE--SMEGWEEB-----ASLETSDHGSDAPQRPPEDEBGAQA 1409
OY 1398 -----CHSLP-----AMPTEPQRYRGR-----GGQCSLYRRVDVYFHTI 1434
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1410 LTAPEGKILLEPCSPIITLDAHELOPA--EGIOAGWQPEASEALEREVENPEFGLEI 1468
OY 1435 EOLTIKHLKMETLAHRELKSLGESSGTTPVGSPFHTAA--RWT---DYSLSF---1486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1469 PE-----GLQWMEBERESEADLDGETLPDSTPLGLYLRSPASPKWDLADGQRSLSPQD 1522
OY 1487 -----PAKEALASDSQNGQEGSCPEEGSDIA---LEDSATDTVAVSPG 1526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1523 AGKEDMGPAVPAAGISGPEEEBEG-----HGSDLSEFEDLTGTEASILPG 1571

```

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972.
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Maris E., Menezes S., Miller N., Nhan N., Pauley A., Peluso D.,
RA Rifken L., Riles L., Tatch A., Trevisan E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.,
RL Submitted (JAN-1993) to the EMBL/Genbank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=96208506; PubMed=8641285;
RA Ozaki K., Tanaka K., Imamura H., Hihara T., Kameyama T.,
RA Nonaka H., Hirano H., Matsura Y., Takai Y.;
RT "Romlp and Rom2p are GDP/GTP exchange proteins (GEPs) for the Rho1p
RT small GTP binding protein in Saccharomyces cerevisiae."
RL EMBO J. 15:2196-2207(1996).
CC -I- FUNCTION: STIMULATES THE EXCHANGE OF RHO1 GDP-BOUND FORM INTO
CC GTP-BOUND FORM.
CC -I- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 019103; AAB67564.1; -.
DR SGD: S0004363; ROM2.
DR InterPro: IPR001180; CNH.
DR InterPro: IPR000591; DEP.
DR InterPro: IPR000219; RhGEF.
DR Pfam: PF00780; CNH; 1.
DR Pfam: PF00610; DEP; 1.
DR Pfam: PF00621; RhGEF; 1.
DR SMART: SM00036; CNH; 1.
DR SMART: SM00049; DEP; 1.
DR SMART: SM00325; RhGEF; 1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.
DR PROSITE: PS50010; DH_2; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 659 846 DH.
FT DOMAIN 252 265 POLY-ASN.
FT DOMAIN 329 336 POLY-HIS.
FT DOMAIN 632 635 POLY-ASP.
SQ SEQUENCE 1356 AA; 152595 MW; 5FBC54211AE7BC92 CRC64;

```

Query Match 3.3%; Score 262.5; DB 1; Length 1356;
Best Local Similarity 20.1%; Pred. No. 6.7e-05;
Matches 200; Conservative 151; Mismatches 311; Indels 331; Gaps 46;

```

OY 235 EADSGIDSGTERFPSTISESLMNRNVLSDPGIDSPQTSFVILARYAONHRRGSDAALLP 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 33 DSDSHGDISQLPRRIENIQNLNVLLSE-----DIANDIT---IAKORRRGVEAA---81
OY 295 LNHGIDGSPKPLITGPEEDYDGYFNN-----SDIIFQDLKLSHPAVLVFLRYIL 349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 82 -----IDSDIP-----NEMKGSNYILISQGTINIKVEPD-----TQSL 116
OY 350 SGADGPGPLFYLCSEVYQOTNPKDSRL-----GQDINMIFLEKAPLRYV-----394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 SSADNTPV-----SSPKKARDATSSHPIVHAKSMSHIYSTSNASQAQAHYNDH 165
OY 395 KPIEMQLQAEIDLRLNNEEDPRNVLCFAOEAVML-----427
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 166 PLPPM-----SPRNEYQKNKSTYAFVPRKKRPSLPQALAGLKKQSSFSFGS 212

```

```

OY 428 -----EIOEO-----INDYRSKRTIGLSLY--- 448
DB 213 ASTTTQARKSPLOGFFGFFSPSSKDLHEOHQHIOHNNINNNNNNTNNGAHYOVG 272
OY 449 GENDLLGDGDPLERRORAEKOLAAGDILSKYEEDRS-----APMDFAVNTFM----- 497
DB 273 SSNSNYPOHSHISIRSMSLNS-STLKNIASSFOSKTSNSRKATQKDYITSNPFDPHH 331
OY 498 -----SHA-----GIRLRESRSCSTAK-----TOSA 519
DB 332 HHHHSSSHSLNNVHSGNSSSVMSGSSNIGLTKTRVSTSLALRKRTSVSGSTLSS 391
OY 520 PDKRWLPFFPKTKKSSNKKER--DALEDKRRNPILRTYIGKPPSSO---SIRPGNV- 573
DB 392 PRSSMTPLASRPVMASSKKPOVYPPAL-----LSRVATKFKSSIQLEHKKDGLVY 444
OY 574 -----RNIIQHPENSHQ-----YDVEPCTORLSTGSEPEDELL 607
DB 445 RDAFTGQAVDVICALIRTSIDRNALLFGRLDAOKLFHDVYEHRLDS---PHEVYE 500
OY 608 -SDSRSRIRLGRGSLKGREEM---KRSKKAENVPRPSDV-----DMDAA 651
DB 501 FTDMS---RFTGTGTVAHDPMLLPNSSSFNSGNSHYPNMGVPSSTSLNSDQATL 556
OY 652 EAARLHOSASSASSLSTRLENPTPT-----PKMGR-----RSIE----- 689
DB 557 TGSRLHMSSSLQOK-NMAAIHNANGVTTLAECSPTCTDALCYSTSCRRLEQOARL 615
OY 690 --SPNLGCTDVIILPHLEDDLGOLSDLEPEPEVOMQHTGVKDVANLTOREIDROQVI 747
DB 616 NLKPGGKLRNISM--ALDDO-----DEEKPSWTSVSKEDWENLPKKEIKROPAI 664
OY 748 NELPTEASHLRLVLDLIFYQRMKRENLM---REELARLPNDELLEIHNMSCEA 803
DB 665 YEVIITENFNFSLTETDTMTKTLAETNIIISADIRKPIKHNVEFHNIDIVSVNRFLKA 724
OY 804 MKKLREBGRIRIDISDPMILARFDPGARELOOVAQFCSY--OSVALLIIRTKORKESR 860
DB 725 LTRDQRSSPVYVRGIGDIYL-RFI-PFFE-----PFSVYASRAYAYALLTETQSVAPY 775
OY 861 FOLEMQEAESHPOCRRLDRLDLIVSEWORLTKYPLLENIITKH--EGGTSEHEKLCSAR 918
DB 776 FARFDDDDMS--SSLRHGIDSEFLSGVSRPGRYMLLVEIKMSTPEDEKSDYEDLSKAM 833
OY 919 DQCEPILKFVNAVQTEBHRLESGYOKRLDATALERASNLAAEFKSLDLT--TRKMIH 976
DB 834 DALRDEMKRIDQASGAADRDHVKLLKOKI-----LFKNRYVNLGLNDERRKIKH 883
OY 977 EGPLTWR-ISK-DKTL--DLQVLLLEDLVLLIQ 1005
DB 884 EGIISRKELSKSDGTVEGDIQFIYLLDNMLFLFK 916

```

```

RT "cDNA cloning of the rat APC gene and assignment to chromosome 18.";
RL Mamm. Genome 6:746-748(1995).
RN [2]
RP MUTAGENESIS.
RC STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;
RX MEDLINE=95148647; PubMed=7846077;
RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
RA Weisburger J.H., Sugimura T., Nagao M.;
RT "Specific 5'-GGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
RT tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATEININ (BY
CC SIMILARITY).
CC -1- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
CC SIMILARITY).
CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D38629; BAA07609.1; -.
CC HSSP; Q02248; 3BCT.
CC InterPro; IPR000225; Armadillo.
CC Pfam; PF00514; Armadillo_seg.6.
CC SMART; SM00185; ARM.5.
CC PROSITE; PS50176; ARM_REPEAT.1.
CC Anti-oncogene; Phosphorylation; Coiled coil; Repeat.
CC KW DOMAIN 1 728
CC FT DOMAIN 1 62
CC FT DOMAIN 125 260
CC FT REPEAT 451 493
CC FT REPEAT 503 545
CC FT REPEAT 546 589
CC FT REPEAT 590 636
CC FT REPEAT 637 681
CC FT REPEAT 682 723
CC FT REPEAT 724 765
CC FT DOMAIN 739 2831
CC FT DOMAIN 1130 1155
CC FT DOMAIN 1356 1575
CC FT DOMAIN 1864 1891
CC FT MUTAGEN 523 523
CC SEQUENCE 2842 AA; 310530 MW; 3CB2EAB8A34EBF47 CRC64;

```

Query Match 3.28; Score 253.5; DB 1; Length 2842;
Best Local Similarity 19.6%; Pred. No. 0.00047;
Matches 358; Conservative 231; Mismatches 717; Indels 517; Gaps 83;

```

OY 8 SIDRASKKOSHLSPIASWLSLSLSDSPERTS---PSHHQPSDT--SENTAGLVQ 62
DB 1217 SAPSSNAKRQSQLH-----PSSAQNRNGQTPKPGTACVPSINDQTMQTYCEVEDPIRFS 1269
OY 63 RCVLIIQKDQHFQFTVSGDRIVLYOVSRPGAA--MKAGVKEGD-----RIIKV 109
DB 1270 RGSLS-----SLSSADEIDGCGDTTQEADASANTLQIAELKENDVYRTSAQDPASDVAV 1323
OY 110 NGTMYTNSHLEVYKLI-----KSGAYALTLGSSPPS----- 143
DB 1324 SOSRTKPSRLQASGLASESARHRAVERFSSGAKSPSKSGAQT-----KSPHYVQDET 1378
OY 144 -----VGVSGIQ--QNSVAGVLRVNP-----IIP-----PP 173
DB 1379 LVFSKCTVSSLSDFESNSTIASVQSEPCSGMVSGIVSPSPLDPSPGQTMPPRSKRTIPP 1438

```


Db 679 RLQAEELLINRVNSTYIVLRHRTKESG 704

RESULT 10

VAV_RAT STANDARD; PRT; 843 AA.

AC P54100;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE VAV proto-oncogene (p95).

GN VAV1 OR VAV.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxId=10116;

RN [1]

RP SEQUENCE FROM N.A.

RA Rivera J., Halem-Smith H.;

RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: PROBABLE EXCHANGE FACTOR FOR A SMALL RAS-LIKE GTP-BINDING PROTEIN. CAN BE ACTIVATED BY TRUNCATION OF THE N-TERMINUS.

CC -1- SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).

CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U39476; AAA98606.1; -.

DR HSSP: P29354; IGR1.

DR InterPro: IPR003247; CH type.

DR InterPro: IPR001715; Calponin_hom.

DR InterPro: IPR002219; DAG_PE-bind.

DR InterPro: IPR001331; GDS_CDC24.

DR InterPro: IPR001849; PH.

DR InterPro: IPR000219; RhogEF.

DR InterPro: IPR000980; SH2.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00307; CH; 1.

DR Pfam: PF00130; DAG_PE-bind; 1.

DR Pfam: PF00169; PH; 1.

DR Pfam: PF00621; RhogEF; 1.

DR Pfam: PF00017; SH2; 1.

DR Pfam: PF00018; SH3; 2.

DR PRINTS: PR00008; DAGPEDOMAIN.

DR PRINTS: PR00452; SH3DOMAIN.

DR PRODOM: PD001527; CH type; 1.

DR SMART: SM00109; C1; 1.

DR SMART: SM00033; CH; 1.

DR SMART: SM00333; PH; 1.

DR SMART: SM00325; RhogEF; 1.

DR SMART: SM00326; SH3; 2.

DR PROSITE: PSS0021; CH; 1.

DR PROSITE: PSS00479; DAG_PE_BIND_DOM_1; 1.

DR PROSITE: PSS0081; DAG_PE_BIND_DOM_2; 1.

DR PROSITE: PSS0010; DH_1; 1.

DR PROSITE: PSS00741; DH_2; 1.

DR PROSITE: PSS0003; PH DOMAIN; 1.

DR PROSITE: PSS0001; SH2; 1.

DR PROSITE: PSS0002; SH3; 2.

KW Proto-oncogene, Phorbol-ester binding, Zinc; SH2 domain; SH3 domain; Guanine-nucleotide releasing factor; Repeat.

FT DOMAIN 1 119 CH.

FT DOMAIN 194 373 DH.

FT DOMAIN 402 504 PH.

FT DOMAIN 516 564 PHORBOL-ESTER AND DAG BINDING.

FT DOMAIN 615 658 SH2.

FT DOMAIN 669 763 SH3.

FT DOMAIN 780 840 SH3 2.

SO SEQUENCE 843 AA; 97953 MW; C4A5CAC045FCB80E CRC64;

Query Match 3.18; Score 245.5; DB 1; Length 843;

Best Local Similarity 21.84; Pred. No. 0.00022;

Matches 143; Conservative 90; Mismatches 254; Indels 169; Gaps 26;

QY 700 ILPLLEDD-IG-----OLSDLEPEPE-----VQNMQHGVKDVYANLQRI-- 741

Db 126 IMPPTEDSALGDEDITSGLSDDQIDTAEEDELDYDCVEN-EEAGEGTEEDLRSESVP 184

QY 742 -----DROEVINELFVTEASHLRTLYLDLIFYQRMKREMLMPREELARLPNLP 792

Db 185 TPPKMTEYDKRCCLREIQOTEETDITGSIQHFMKPLQR--FLKPDOMETFVINEE 242

QY 793 LIETHNSWCEAMKRLREBPILRIDISPMILARDGPAREELQ-----VAAQFCS 842

Db 243 LLSVTHFR---LKLKLD-----ALSGPATMLYGVFIYKKERFLVYGRYCS 285

QY 843 YQSVALELIRTKOKESRFOLFQEAESHPCRRLOGLDIVESMORLTFTYPLLENIIR 902

Db 286 QVESAIITHLDQVATAREDVQMKLEECGQRANNGHFTIRDLAMPQWVLYLTHLLDELVK 345

QY 903 HTEGTSSEHEKLCRAPDQCREILKFVNEAVKOTENRRHLEGYQ---RLDATALERASNP 959

Db 346 HTQ-DTTEKENLRALDAMDRLAQCVNEVRKNETLQIINFQLSIENLDQSLANVGRPK 404

QY 960 LAEFKSLDUTTKRMHIEGPITRKISDKTLDQVLLLEDVLYLQOEERLLKCKSKT 1019

Db 405 IDGELKITVSERR-----SKT-DRYAFLLDALALKCKRSGSYDJKAS-- 446

QY 1020 AVGSDDSKQTFSPYLKINAVLIRSATDKR-----AFLTIGSELPQIYELVLT 1072

Db 447 -----VNLSFOVRDSSGERDNKKMSMFLII--EEOGAQGIELFPKTR 489

QY 1073 SDKNIMELLEAVONATKRGAPIPIHSPSGQSPA-----YOG----- 1114

Db 490 ELKKKMEQEFMAISNI--YPENATANGHDFQMFSEFTTSCRAQMLLGTFTYQGRYCY 547

QY 1115 -----STSSRYEINDSEVYHTEKPKKLPGSPGPQYQVQKOLTAQGPVQOEDEEE 1166

Db 548 RCRAPAHKECLGRVPPGCRDQFSGTMRKRL-----HRAADKKRNEELIPKREVCQ 601

QY 1167 LRTLPRAAPSIDGENRGIRTRDPVLLALGTPLMEGLADAA---LEDVENLRHLILMSLT 1223

Db 602 YGIRP--PPGAFG-----PFLRLNPGDIYELTAELAEHRTWMEGRNATNEVGH--F 648

QY 1224 PGHTVKTQAAGEPEDDLTPPSVVSITSHPW-----DPSPGQAPTISDSTRLAR 1273

Db 649 PCNRVRYVHGPPQD-----LSVHLMYAGPMERAGAGILTRNSDGTLYLR 694

RESULT 11

VAV3_MOUSE

ID VAV3_MOUSE STANDARD; PRT; 847 AA.

AC G9ROC8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE VAV-3 protein.

GN VAV3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;

RN [1]

```

RP SEQUENCE FROM N.A.
RA Trenkle T., McCelland M., Welsh J.:
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: EXCHANGE FACTOR FOR GTP-BINDING PROTEINS RHOA, RHOG AND,
CC TO A LESSER EXTENT, RAC-1. BINDS PHYSICALLY TO THE NUCLEOTIDE-FREE
CC STATES OF THOSE GTPASES (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, ALPHA (SHOWN HERE) AND BETA;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 CALPONTIN-HOMOLOGY (CH) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DEL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/csb.ch).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF067816; AAF09171.1; -
DR HSSP: P12931; 1A1E.
DR MGD: MGI:1888518; Vav3.
DR InterPro: IPR003247; CH_type.
DR InterPro: IPR001715; Calponin_hom.
DR InterPro: IPR002219; DAG_Pe_bind.
DR InterPro: IPR001331; GDS_CDCC24.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhGEF.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR003096; SM22_calponin.
DR Pfam: PF00307; CH; 1.
DR Pfam: PF00130; DAG_Pe_bind; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhGEF; 1.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRINTS: PR00888; SM22CALPONIN.
DR PRODOM: PD001527; CH_type; 1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00233; CH; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhGEF; 1.
DR SMART: SM00252; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS50021; CH; 1.
DR PROSITE: PS00479; DAG_Pe_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_Pe_BIND_DOM_2; 1.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS00741; DH_1; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 2.
KW Phorbol-ester binding; Zinc; SH2 domain; SH3 domain; Repeat;
KW Guanine-nucleotide releasing
FT DOMAIN 1 119 CH.
FT DOMAIN 192 371 PH.
FT DOMAIN 400 502 DH.
FT DOMAIN 514 562 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 592 660 SH3 1.
FT DOMAIN 672 766 SH2.
FT DOMAIN 788 847 SH3 2.
SO SEQUENCE 847 AA; 97946 MW; 9A6B63F0D9E60F8F CRC64;

```

[2]
RN SEQUENCE OF 1-93 FROM N.A.
RP MEDLINE=91172176; Pubmed=2005887;
RA Ketzav S, Cleveland J.L., Heslop H.E., Pulido D.;
RT "Loss of the amino-terminal helix-loop-helix domain of the vav proto-
oncoprotein activates its transforming potential.";
RL Mol. Cell. Biol. 11:1912-1920(1991).
CC - FUNCTION: PROBABLE EXCHANGE FACTOR FOR A SMALL RAS-LIKE GTP-
BINDING PROTEIN. CAN BE ACTIVATED BY TRUNCATION OF THE N-TERMINUS.
CC - TISSUE SPECIFICITY: WIDELY EXPRESSED IN HEMATOPOIETIC CELLS BUT
CC NOT IN OTHER CELL TYPES.
CC - SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.
CC - SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
CC - SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
BINDING DOMAIN.
CC - SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC - SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64361; CAA45713.1; -.
DR EMBL: M59833; AAA63402.1; -.
DR PIR: A39576; A39576.
DR PIR: S36941; S36941.
DR HSSP: P29354; 1GRI.
DR TRANSFAC: T01230; -.
DR MGD: MGI:98923; Vav.
DR InterPro: IPR003247; CH_type.
DR InterPro: IPR001715; Calponin_hom.
DR InterPro: IPR002219; DAG_pe_bind.
DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00307; CH; 1.
DR Pfam: PF00130; DAG_pe_bind; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00018; SH3; 2.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD001527; CH_type; 1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00033; CH; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhoGEF; 1.
DR SMART: SM00252; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS50021; CH; 1.
DR PROSITE: PS00479; DAG_pe_bind_dom_1; 1.
DR PROSITE: PS50081; DAG_pe_bind_dom_2; 1.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS00741; DH_1; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 2.
KW Proto-oncogene; Phorbol-ester binding; Zinc; SH2 domain; SH3 domain;
KW Guanine-nucleotide releasing factor; Repeat.
FT DOMAIN 1 119 CH.
FT DOMAIN 194 373 DH.
FT DOMAIN 402 504 PH.
FT DOMAIN 516 564 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 617 660 SH3 1.
FT DOMAIN 671 765 SH2.

FT DOMAIN 782 842 SH3 2.
FT CONFLICT 29 29 0 -> E (IN REF. 2).
SQ SEQUENCE 845 AA; 98136 MM; 3666DCDD1C5225DA CRC64;

Query Match 3.0%; Score 240; DB 1; Length 845;
Best Local Similarity 21.4%; Pred. No. 0.0004;
Matches 168; Conservative 93; Mismatches 300; Indels 224; Gaps 32;

QY 700 ILPHLEDDL-----GQLSDLEPRPE-----VQWQITVCKDVYANLTQREI-- 741
DB 126 IMPPTEDSALNDEDIYSGLSQIDIDTAEDDEDLYDCVEN-EBAEGDEIYEDLMRESVP 184
QY 742 -----DROEYINLEFTEASHLFTLVLLIFQYRKREMLPREELARFLPNLPE 792
DB 185 TPQKMTYDKRCCCLRIQOOTEKTYTTLQSIQHFKEPLQR--FLKQDMETIFVNIIE 242
QY 793 LIEIHNSCEAMKRLREBPRIINDISDPMLARDPAREELOQ-----VAAQFCS 842
DB 243 LFSVHTHF-----VNLHSEFOVRDSSGGERDNKKSHMFLI--EDQAGQYELFEFKT 488
QY 843 -YQSVALELIRTKORKESRQLFMQEAESHPCRRQLQRLIVSEMQRLTKYLLLENTI 901
DB 286 QVESASAKHLQVATYARD-VQMKLECSQRANNGRFTLRDLVWPQRYLVKYLHLLQELV 344
QY 902 KHEGGSSEHEKLCARDOCREILKFNVEAVKQTEENHRELEGO---KRLDATALERASN 958
DB 345 KHTQDAT-EKENRLALDAMRDLAQCVEYKKNETLRQTLTNQSLTENLDQSLANTGRP 403
QY 959 PLAAEFKSLDLTTRKMIHEGRLTWIRSKDTLDLOVLLLEDLVLLQROBERLLKCHSK 1018
DB 404 KIDGELKITSVERR-----SKT-DRYAFLLDKALLICKRRGSDYDLMA-- 446
QY 1019 TAVGSSDSKQTFSPVLKLNVLRSVATDKR-----AFFITCTSELGRPYLEYALT 1071
DB 447 -----VNLHSEFOVRDSSGGERDNKKSHMFLI--EDQAGQYELFEFKT 488
QY 1072 SSDKNIMELLEBAVQNAATKHPGAAPRIHPSPGSOEPA-----YQGSTS 1117
DB 489 RELKKMMEDQFEMALISNI--YPENATANGHPDMFSEETTSCKACOMLLRGTFYGYCR 546
QY 1118 SRVEI-----NDSEVYHTEKEPKRLDEGPGPEGRVODKQLIAGGEVQEE 1162
DB 547 YCRAPAHKCELCGRVPCGRHGFAGTMMKDKL-----HRRADQKKNELGLPKMEV 599
QY 1163 DEEELRTLPRAPPSLOENGCITRQDVLLALGCLIMEGLAQA---LEDVENLKHLL 1219
DB 600 FOEYVGIRP--PPGAFG-----PFLRLNPGDIVELTKAEAEHNMWCGRTATNEVG 648
QY 1220 WSLPLGHTVTKQAAGEBEDDLPTPSVYSITSHPW-----DQSPGQAPTISDSTRLAR 1273
DB 649 W--FPCNRVHPYVHGPPQD-----LSVLMTAGPMEERAGACGILLNRSDDGYTLVR 696
QY 1274 -----PEGQPEGEDEVAVSSLAHLPPRTSSGCVMS-DE 1306
DB 697 QRVKDAEFAISIKYVNEVYKIMISEGLYRTTEKKAFFGLLELVFQGNLSKQCFPS 756
QY 1307 LD-----NPPAAEAASTPPAASYKVKVSLIPGGVGGAATVAAASMALIPDSGQSESEL 1359
DB 757 LDYTLFPYKPEPRRAISKPPAGSTKY-----FQTAKARYDFPCARD---RSEL 801
QY 1360 SEVEG 1364
DB 802 SLKEG 806

RESULT 13
TRIO HUMAN
ID TRIO HUMAN STANDARD: PRT; 3038 AA.
AC 075962; 013458;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)

FT MTGAGEN 1372 1372 K->A: LOSS OF NUCLEOTIDE EXCHANGE
 FT MTGAGEN 1375 1375 L->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE
 FT MTGAGEN 1378 1378 ACTIVITY.
 FT MTGAGEN 1379 1379 K->A: NO CHANGE IN NUCLEOTIDE EXCHANGE
 FT MTGAGEN 1379 1379 ACTIVITY.
 FT MTGAGEN 1379 1379 E->A: 30% DECREASE IN NUCLEOTIDE EXCHANGE
 FT MTGAGEN 1379 1379 ACTIVITY.
 SQ SEQUENCE 3038 AA; 341611 MW; 28620F3B513EB74B CRC64;

Query Match 3.0%; Score 238.5; DB 1; Length 3038;
 Best Local Similarity 17.6%; Pred. No. 0.0025;
 Matches 272; Conservative 195; Mismatches 478; Indels 599; Gaps 66;

QY 229 LVSQADSG--LDSGTERPSSISESLMNRNSVLSPGGLDSPOTSVYLLARVAQHR-- 284
 DB 1189 ISSDSNRSSKSLQDIIIPASIPGSEVKLRDAHLENEKRSARKEFIIMAEIYQEKAY 1248
 QY 285 ---RQSDAALLPLNHQIDSP---KPLIT--GPEEDYDGYFNNESEDIFODLEK 333
 DB 1249 VRDLRECDMTYLEMNT--SGVEIPIPGIYNKELLIFGNMGEIYE---FHN--NIFLELEK 1302
 QY 334 LKSHPA-----YLVFLRYILSQADPGPLFYLCSEVYQGTNPKDSRSLGKDIW 382
 DB 1303 YEQLPEVDVGHCFVTWADRFQMYVTYCKNKPDSQIILEHAGSYDEIQOR--HGLANSIS 1360
 QY 383 NIFLEKNAPL-RVKIPIEMLAELRLRNEDPRNVLCENQAEVWLEQOENINYSRKRT 441
 DB 1361 SYLIK---PVORITKYQLLKEKEL--LTCCBEGKEITKDGLE--VMSVPRKAPN-----A 1408
 QY 442 LGISLGVENDLGLDG-----DPL-----REROMAEKQALALDILSKYEED 484
 DB 1409 MHLISLMBGFDENISQGLDIIQESFOYWDPKTLIRKGERHLF--LFEMSLVFSKEVD 1465
 QY 485 RSAPMDAVNT--FMSHAGITLRESRSSC-----TAE-----KTOSAPDKMWLP 527
 DB 1466 SSGRSKYLKSKLPTSELGVEHEHGDPCRFALWVGRTPTSDNKIVLKASIEKODMIK 1525
 QY 528 FF-----PKTKSSNSKKEKDALEDKRNP----- 553
 DB 1526 HIREVIERTHLKGALKEPHIHTKAPATROKGRDGEDDSOGDDSDPDTIISART 1585
 QY 554 -----LRYIGKPKSSSQSIKPGVNRNIIQHFNESHQYDVE----- 590
 DB 1586 SQNTLSDSKLSGGCELIVVINDFTACNSNELTIRGGTVEYLE-----RPADKDKWCLVR 1640
 QY 591 -----PGTORL-----STGFPEDLLE 607
 DB 1641 TTDRSPAAGLIVPCGSLCIAHSRSSMEMEGIFNHNKDSLVSNDASPPASVSLQPHMIG 1700
 QY 608 SDDS-----RSEIRLGRSGSLKGR-----EEMKRSRKANVPRPSVDYMD 648
 DB 1701 AQSPPGKRRGENTLRKMLTSPVRLSSGKADGHYKKAHKKRSREVRKSDAGSQDSD 1760
 QY 649 AAA-----EAARLHQASASSASS----- 666
 DB 1761 DSAATPODEVEERGRNEGILSSGTLSSSSGMSGCEBEGEGADAVPLPPAIQOHS 1820
 QY 667 -----LSTRSLNPPPTPTP-----KMGKRSLIESPLGCTIVYL 701
 DB 1821 LLOPDSODDKRASSRLVPRPSSPTPSAAELVSAIEELVKSMALEDPRSSILVDOGSSS 1880
 QY 702 PHLEDDLGOLSDLEPEREYVOMOHITGKDVANILTOREIDROVINELEFTEASHLRTL 761
 DB 1881 PSFNPDSNLSLSSSPIDEMERKSSSLK-----RRHVVLDLVELTENDYVRLD 1929
 QY 762 RVLDLIFYQRRKKNEMPREELAR--LFPNLPPELLIETHNSWCAMKLRREGPIIDIS 818
 DB 1930 GYV--VEGYMALMKEDGVDDMKGDKIVFGNIHQIDMHRDF-----LGELEKCLE 1980
 QY 819 DPMIARDEGARRELOOYVAOFCSYQVALELITKQKRSERFOLPMQEAESHQC----- 874

DB 1981 DP-----EKLGSLEFYKH-----ERRLHMYIAYCONKPKSEHIY 2013
 QY 875 -----RRLDRLIYSEMORIKRYPILLENIITKHGSGSEHKKLCORAR 918
 DB 2014 SEYIDFEEDLKORLGHRLDITDLIRPVORIMKYOLLDLDPKYSKASLDISELERA- 2072
 QY 919 DQCEILKFVNEAVKQTEENHRRLEGYOKRDATALEBASNPAAEFKSLDITRKMIEG 978
 DB 2073 ---VEVMCIYPRKRDMMANVGRLOGFD-----GKIYAOG 2103
 QY 979 PLTWIRISKRTLLDLYLLLEDVLLOROEERLLKCHSK-----TAVGSS--DSKOT 1029
 DB 2104 K-----LLLOD--FLVLYDQDAGLLPRCERBRIFLEQIVFSPBLDKKG 2147
 QY 1030 FS-----PVKLKNAV-LIRSVATDKRAFFICTSFLSPQIYELVALTSSDKNT---W 1078
 DB 2148 FSNPGLFKNSIKVSCLEENENEDCPKAL--TSHTG--DVEVFILHSSSPSVKQIW 2203
 QY 1079 MELLEAVONATKHPGAAPIPH-----PSPPG---SQ 1108
 DB 2204 IHEINOLLENOBNFLNLTSPIEYQRNHSGGGGSGSGSGSGGSGGAPSGGSGHSG 2263
 QY 1109 EPAYOGSTSSRVEIINDESVYHTEKEPKKLPEGPEPEODVODKOLLAOGEVQEBDEELR 1168
 DB 2264 GPSSCGGAPST-----SRSRPSRIPQ----- 2284
 QY 1169 TLPAPPSLDGENKRTDPVLLALTGPLMBSIADALEDEVENLHILMSLLPGHTV 1228
 DB 2285 -----PVNHPHRYL-----V 2294
 QY 1229 KTOAGEPEDD-----LTPPSVVSITSHPDPSGPGQARTISDSTRLARPEGSSQPEGD 1283
 DB 2295 SSAASQAEADKMSGTSTPGPSTL-----PPGCAPEAGPSAPSRRPADAGESE 2344
 QY 1284 VAVSIAHLHPRTSSGVMPSPELDRNPAEASTEPAAATYKYVR--KVSLLPQG--GVGA 1340
 DB 2345 REAEPV---PKMK---VLESP---RKGAAVSSSSDPAKAKARASLTGLTLPGLKPRAGA 2394
 QY 1341 AKVAG---SNAIPDSGQS---ESELSEVEGAQATGNCFYVSPAGPL-----D 1383
 DB 2395 ASPLNLSGAVPSLGEKPEPPSPLOK-----GGSFWSSIPASPARPSGFTPPGD 2446
 QY 1384 STEPTGTPPSPSQCHSLPAMPTEPQRYKGVKGCCSSLVRRDV 1427
 DB 2447 SDSLRQRTPR-----HAAPGKQTD-----RMSCTSSASEQSV 2478

RESULT 14
 ECT2_MOUSE
 ID ECT2_MOUSE STANDARD; PRT; 738 AA.
 AC 007139;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ECT2 protein (Epithelial cell transforming sequence 2 oncogene).
 GN ECT2.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=93218723; Pubmed=8464478;
 RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;
 RT "Oncogene ect2 is related to regulators of small GTP-binding
 RT proteins";
 RL Nature 362:462-465(1993).
 RN [2]
 RN ERRATUM.
 RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;
 RL Nature 364:737-737(1993)
 CC -!- FUNCTION: BINDS HIGHLY SPECIFICALLY TO RHQ, RHOC AND RAC
 CC PROTEINS, BUT DOES NOT APPEAR TO CATALYZE GUANINE NUCLEOTIDE

```

CC EXCHANGE. TRUNCATION OF THE N-TERMINUS PUTS THE PROTEIN IN AN
CC ACTIVATED STATE.
CC -I- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS. ALSO DETECTABLE
CC IN KIDNEY, LIVER AND SPLEEN.
CC -I- SIMILARITY: CONTAINS 2 BRCT DOMAINS.
CC -I- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -I- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L11316; AAA37536.1; -.
CC MGD; MGI:95281; Ect2.
CC InterPro: IPR001357; BRCT.
CC InterPro: IPR001331; GDS_CDC24.
CC InterPro: IPR001849; PH.
CC InterPro: IPR000219; RhogEF.
CC Pfam; PF00533; BRCT; 2.
CC Pfam; PF00621; RhogEF; 1.
CC SMART; SM00292; BRCT; 2.
CC SMART; SM00233; PH; 1.
CC SMART; SM00325; RhogEF; 1.
CC PROSITE; PS50172; BRCT; 2.
CC PROSITE; PS50010; DH_2; 1.
CC PROSITE; PS00741; DH_1; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
CC Guanine-nucleotide releasing factor; Repeat; Proto-oncogene.
CC FT DOMAIN 1 76 BRCT 1.
CC FT DOMAIN 91 179 BRCT 2.
CC FT DOMAIN 277 466 DH.
CC FT DOMAIN 500 619 PH.
CC SEQUENCE 738 AA; 83685 MW; 3D270CBF1570C681 CRC64;

```

Query Match 3.0%; Score 237; DB 1; Length 738;
 Best Local Similarity 19.4%; Pred. No. 0.00046;
 Matches 133; Conservative 118; Mismatches 259; Indels 174; Gaps 25;

```

QY 645 VMDA-AAEARLHQA-----SSASLSL-----RSLNPPTPT 680
DB 166 IQMDADAGETMYLYERKANTPELKSVSLSTPNSNRKRLKETLAOLSRPTDISPP 225
QY 681 PKMGRSIESPNLGFCTDYILPHLEDDIGOLSDLEPEPEVQNMQHTVGKDVAN----- 735
DB 226 PKR-RSAE-----HSL--SIGSLDISNTEPSSIHGFTPKSCAKSSRSST 269
QY 736 -LTQREIDROEVINELFVTEASHLRLRVLDLIFYQRMKEN-----LMPREELARLPN 789
DB 270 PVPKQASAKWQAKELYQTESNVNLTATIQFVPLEEGRGCPILAPEIKTIFGS 329
QY 790 LDELIEIHNWCEAMKKLKEEGPIIDISDPMIARPDGA-----RELQOAAAP 840
DB 330 IDPIEFVH-----MK-----IKDLEDLIANDESRSIGDIFLYAKADLVKYPF 375
QY 841 CSYQSALELLIRTKQKESRFQFQMEASHPOCRRLQRLDIVSEMRLLTKYPLLENI 900
DB 376 VAFPEKSKMIKCEKQKRFHAFKLNQAKPECGOSLVELLIRVOQLPSALLIND 435
QY 901 IKHTEGTSHEHKLRCARDQREILKFNVEAVKQTENRHRLLEGYOKRLDATALERASNDL 960
DB 436 KKHTEADENPKSTLEKAIISLKEVMTHINEDKRTKQAQKQIDVYEVDC-----PANLL 491
QY 961 AAEFSLDITTKMHHEGLTWIRISKDKTLDAQVLLLELVVLQROEERL--LAKCHSK 1018
DB 492 SSRLSYORVETVSLGEHNC-----DREQOYTLFNFQCLIEARRHNRVIGTFSPHQR 545
QY 1019 TVAGSSDSKQTFSPYLKNAVL--IRSVANDKRAFLICTSELGPPQIYELVALTSSD--K 1075

```

```

DB 546 TRPPASLKIHIMPLSQIKKVLDIRTEDCHNAPFALLVRPPEQANVLSFOKTSSELPK 605
QY 1076 NIWMELEEAQVATKHPGAPPIHPSPGSGEPAYOGSTSSRYEINDSEVYHTEKEPK 1135
DB 606 ESKLMLKCHRVANTICKAENL--MYVADPE-----FEVNTKDM----- 644
QY 1136 KLEPGGPEQRVODKOLINQGEPEDEBEL--TLPPAPSLDGENGIRTRDPVLLA 1193
DB 645 -----DSTLSRASRAIKTKSKVYRAFESKTP-----KRALR-----MA 679
QY 1194 LFGPLMEG-----LDAALAEDEVNRLHILMSLPGHVFYKTOAGEPEDDLTPPS 1245
DB 680 LSSHSSEGRSPSSCKLAVSLRSTSSL-----AQPSPSLVSLPS 721
QY 1246 VVSTSHPMWDPGSPGQAPTIIDST 1269
DB 722 FFERRSH-----TLRST 734

```

RESULT 15

```

ID VAV_HUMAN STANDARD; PRT; 845 AA.
AC P15498; Q15860;
DT 01-APR-1990 (Rel. 14; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 01-MAR-2002 (Rel. 41; Last annotation update)
DE VAV proto-oncogene.
GN VAV1 OR VAV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Denkiner D.J., Borges C.R., Shaw C.L., Cushman A.M., Kawahara R.S.;
RT "Transcriptional regulation of the vav proto-oncogene."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 62-845 FROM N.A.
RA MEDLINE=9005432; PubMed=2477241;
RA Katay S., Martin-Zanca D., Barbacid M.;
RT "vav, a novel human oncogene derived from a locus ubiquitously
RT expressed in hematopoietic cells."
RL EMBO J. 8:2283-2290(1989).
RN [3]
RP SEQUENCE OF 1-61 FROM N.A.
RA MEDLINE=91172176; PubMed=2005887;
RA Katay S., Cleveland J.L., Heslop H.E., Pulido D.;
RT "Loss of the amino-terminal helix-loop-helix domain of the vav proto-
RT oncogene activates its transforming potential."
RL Mol. Cell. Biol. 11:1912-1920(1991).
RN [4]
RP SEQUENCE OF 299-837 FROM N.A.
RA Romero F.;
RT Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 299-334 FROM N.A.
RA MEDLINE=96038895; PubMed=7478592;
RA Ramos-Morales F., Romero F., Schweighoffer F., Bismuth G., Camonis J.,
RA Tortolero M., Fischer S.;
RT "The proline-rich region of Vav binds to Grb2 and Grb3-3."
RL Oncogene 11:1665-1669(1995).
RN [6]
RP SIMILARITY TO CDC24 FAMILY.
RA MEDLINE=92228488; PubMed=1565462;
RA Adams J.M., Houston H., Allen J., Lints T., Harvey R.;
RT "The hematopoietically expressed vav proto-oncogene shares homology
RT with the dbl GTP-GDP exchange factor, the bcr gene and a yeast gene
RT (CDC24) involved in cytoskeletal organization."
RL Oncogene 7:611-618(1992).
CC -I- FUNCTION: PROBABLE EXCHANGE FACTOR FOR A SMALL RAS-LIKE GTP-
CC BINDING PROTEIN. CAN BE ACTIVATED BY TRUNCATION OF THE N-TERMINUS.
CC -I- TISSUE SPECIFICITY: WIDELY EXPRESSED IN HEMATOPOIETIC CELLS BUT

```

CC NOT IN OTHER CELL TYPES.
CC -1 SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.
CC -1 SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1 SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1 SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAIN.
CC -1 SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1 SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -1 CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 322 TO 355 DUE TO A FRAMESHIFT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF030227; AAC25011.1; -
DR EMBL; AF030201; AAC25011.1; JOINED.
DR EMBL; AF030202; AAC25011.1; JOINED.
DR EMBL; AF030203; AAC25011.1; JOINED.
DR EMBL; AF030204; AAC25011.1; JOINED.
DR EMBL; AF030205; AAC25011.1; JOINED.
DR EMBL; AF030206; AAC25011.1; JOINED.
DR EMBL; AF030207; AAC25011.1; JOINED.
DR EMBL; AF030208; AAC25011.1; JOINED.
DR EMBL; AF030209; AAC25011.1; JOINED.
DR EMBL; AF030210; AAC25011.1; JOINED.
DR EMBL; AF030211; AAC25011.1; JOINED.
DR EMBL; AF030212; AAC25011.1; JOINED.
DR EMBL; AF030213; AAC25011.1; JOINED.
DR EMBL; AF030214; AAC25011.1; JOINED.
DR EMBL; AF030215; AAC25011.1; JOINED.
DR EMBL; AF030216; AAC25011.1; JOINED.
DR EMBL; AF030217; AAC25011.1; JOINED.
DR EMBL; AF030218; AAC25011.1; JOINED.
DR EMBL; AF030219; AAC25011.1; JOINED.
DR EMBL; AF030220; AAC25011.1; JOINED.
DR EMBL; AF030221; AAC25011.1; JOINED.
DR EMBL; AF030222; AAC25011.1; JOINED.
DR EMBL; AF030223; AAC25011.1; JOINED.
DR EMBL; AF030224; AAC25011.1; JOINED.
DR EMBL; AF030225; AAC25011.1; JOINED.
DR EMBL; AF030226; AAC25011.1; JOINED.
DR EMBL; X16316; CAA34383.1; ALT_FRAME.
DR EMBL; M59834; AAG3267.1; -
DR EMBL; X83931; CAA58783.1; -
DR PIR; S05382; TVHUV.
DR PIR; B39576; B39576.
DR HSSP; P29354; IGR1.
DR TRANSFAC; T00880; -
DR MIM; 164875; -
DR InterPro; IPR003247; CH_type.
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR002219; DAG_PE_bind.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhogEF.
DR InterPro; IPR000960; SH2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR003096; SM22_calponin.
DR Pfam; PF00307; CH; 1.
DR Pfam; PF00130; DAG_PE_bind; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhogEF; 1.
DR Pfam; PF00017; SH2; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRINTS; PR00888; SM22CALPONIN.
DR ProDom; PD001527; CH_type; 1.

DR SMART; SM00109; C1; 1.
DR SMART; SM00033; CH; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhogEF; 1.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50021; CH; 1.
DR PROSITE; PS500479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS00741; DH_1; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 2.
KW Proto-oncogene; Phorbol-ester binding; Zinc; SH2 domain; SH3 domain; Guanine nucleotide releasing factor; Repeat.
FT DOMAIN 1 119 CH.
FT DOMAIN 194 373 DH.
FT DOMAIN 402 504 PH.
FT DOMAIN 516 564 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 617 660 SH3 1.
FT DOMAIN 671 765 SH2.
FT DOMAIN 782 842 SH3 2.
FT CONFLICT 264 264 SH3 2.
FT CONFLICT 718 718 A -> P (IN REF. 2).
FT CONFLICT 718 718 I -> TV (IN REF. 2).
SQ SEQUENCE 845 AA; 98313 MW; AC3BC9736FD2F138 CRC64;

Query Match 3.0%; Score 233.5; DB 1; Length 845;

Best Local Similarity 21.7%; Pred. No. 0.00079; Matches 151; Conservative 93; Mismatches 267; Indels 185; Gaps 30;

QY 711 QLSDELPEPE-----YQNMCHTVGKDVANLTQRE-----IDROEVINEFEVTE 754
DB 148 QIDDIYEDDEDLYDCYBN-EEAGDEIYEDIMRSEVSNPPKTEYDKRCCCLREIQOTE 206
QY 755 ASHLRTLVLDLIFYQMRKKNLMPREELARLPNLPETLEIHNWCAMKRLREGPII 814
DB 207 EKYTDITGSIQGHFKRPLQR-FLKPDIDIEIFINIEDLLRVHTHFKEMKE----- 256
QY 815 RQISDMLARFDSAPARELQO-----VAAGCS-QYSVALLLIRTKOKRESFOL 863
DB 257 -----ALGTPGANLYOVPIKRYKEREFLVYGRCSOVESASKHLDVYAAARE-VQM 306
QY 864 FMOEASHPOCRLOLRDLIVSEMOURLTKYLLLENIIKHTEGSGSEHEKLCARPOCRE 923
DB 307 KLECCSQRRANNGRFTLRDLIMVQRYVLYKHLLOELVHTQ-EEAKKEKRLALDAMD 365
QY 924 ILKEVNEAVKQTEENRRHRELEQY--KRLDATALERASNPPLAEFKSLDITRRMHEGPI 980
DB 366 LAQCVNEVKRDNETLRQITNFQSLIENLDQSLAHYGRPKIDGLK--ITS----- 413
QY 981 TWRISKDTLDLVLLLEDLVVLQROEERLLK-----CHS-----KTAVGSSSKOTFSP 1032
DB 414 ---VERRSKMDRYAFLDLNALLLCKRRGSDYDLKDFVNLHSFOVRDSSGDRNK----- 465
QY 1033 VLKLNVLILRSVATDRAEPIICTSELGPOIYELVALTSSDPKNIMLEELPEAVQATKH 1092
DB 466 -----KSHMFLIIEIDQ-AGQTELEFKRELKKKMKDEQEMALISNI--Y 507
QY 1093 PGAAPPIPIHSPPGSQEPA-----YQGSTSRVEINDSEVYHTEKEPKKLP 1138
DB 508 PENATANGHDQFMFSFEETTSCKACQMLRGTFYQGYRCHRCAS----AHKECLGRVPP 563
QY 1139 EG-----FGP-----EQRVQDKQLAAGEPYQDEDEELKRLPPAPPSLGENNGIRT 1186
DB 564 CGRHGQDFPGTKMKDKLHRAADQKKRNEGLPRMEVFOEYEGLP--PGAIG----- 614
QY 1187 RQVLLALGLPLMBGLAA--LEDVENLRLLILMSLLPGHTVYTOAAGEPEDDLPFT 1243
DB 615 --PFLRLNPGDVLVETLKAEBQNMWEGKNTSTNIEGW--PPCRKRVAPYVHGPPD----- 665
QY 1244 PSVASTISHPWDPGSPGOAPTISDSTRLARPEGSQEGEDVAVSSLAHVPPTRSSGVD 1303

Db 666 -----LSVHLMWYAG-----PMERAGAESILAN-----RSDGTF- 693
Oy 1304 SPELDRNPAAEASTEPASYKV-VRKVSILPBGGV 1338
Db 694 ---LVROKVKDAAEFASIKYNEVYKHKIMTAEG 726

Search completed: September 18, 2002, 10:43:27
Job time: 328 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 ; Search time 31.34 Seconds
(without alignments)
4681.826 Million cell updates/sec

Title: US-09-695-795-4

Perfect score: 7883

Sequence: 1 MSIRLPHSIDRSASRKQSHL.....EGSDIALEDSATDTAVSPGP 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_71:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922.5	11.7	2559	2 T09144	probable guanine n
2	672	8.5	979	2 A89633	protein F13B6.6 [1
3	521	6.6	1693	2 T30867	Rho-guanine nucle
4	462	5.9	573	2 T49342	hypothetical onco
5	428	5.4	424	2 I38434	p47 LBC oncogene -
6	318	4.0	1805	2 A34736	nestin - rat
7	297.5	3.8	1737	2 A59235	unconventional myo
8	297	3.8	1804	2 T34518	nestin - golden ha
9	262.5	3.3	1356	2 S51389	ROM2 protein - yea
10	261	3.3	1158	2 T50454	probable rho1 GDP-
11	260.5	3.3	1334	2 T41524	rho1 gdp-gtp excha
12	256.5	3.3	1046	2 T34566	hypothetical prote
13	255.5	3.2	5327	2 T13564	microtubule-associ
14	247.5	3.1	1618	2 S21424	nestin - human
15	239.5	3.0	1045	2 JC5795	CDEP protein - hum
16	239	3.0	2101	2 A42184	nuclear mitotic ap
17	237	3.0	738	2 S32372	transforming prote
18	231.5	2.9	460	2 G01210	guanine nucleotide
19	229	2.9	2845	2 I49505	adenomatous polypo
20	228.5	2.9	1829	2 T21635	hypothetical prote
21	227	2.9	878	2 T51940	gene VAN2 protein
22	226	2.9	2774	2 A43359	microtubule-associ
23	226	2.9	1591	2 A54146	invasion-inducing
24	224.5	2.8	648	2 T33339	hypothetical prote
25	223.5	2.8	1155	2 S64365	GDP/GTP exchange p
26	222	2.8	2954	2 T14156	kinesin-related pr
27	221.5	2.8	519	2 T38402	guanine nucleotide
28	221	2.8	1252	2 T14272	coractin-binding
29	220.5	2.8	1621	2 A82255	hypothetical prote

30	220.5	2.8	1684	2 JN0057	gravin - human
31	219	2.8	2044	2 T13704	still life protein
32	219	2.8	2064	2 T13707	still life protein
33	215	2.7	875	2 T19678	hypothetical prote
34	213.5	2.7	961	2 A55380	faciogenital dyspl
35	213.5	2.7	1547	2 T28657	blackjack protein,
36	212	2.7	3938	2 T42761	Bassoon protein -
37	211.5	2.7	753	2 S62411	probable guanine n
38	211.5	2.7	1427	2 S22695	restin - human
39	211.5	2.7	1791	2 T02345	hypothetical prote
40	209	2.7	1131	2 T15617	hypothetical prote
41	209	2.7	1275	2 T41523	hypothetical rho1
42	207.5	2.6	3942	2 T42730	Bassoon protein -
43	206.5	2.6	2442	2 T08621	centrosome associa
44	206.5	2.6	2715	2 T13049	eyelid - fruit fly
45	204.5	2.6	862	2 T34342	hypothetical prote

ALIGNMENTS

RESULT 1

T09144
N:probable guanine nucleotide exchange factor RHOGEF2 - fruit fly (Drosophila melanogas
C:Species: Drosophila melanogaster
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Nov-2000
C/Accession: T09144; T09223
R/Recker, U.; Perrimon, N.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z16586
A:Accession: T09144
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12559 <HAES>
A:Cross-references: EMBL:AF031930; NID:g2687355; PID:g2687356
R/Barrett, K.; Lepin, M.; Settleman, J.
Cell 91, 905-915, 1997
A>Title: The Rho GTPase and a putative RHOGEF mediate a signaling pathway for the cel
A:Reference number: Z16618; MID:98088790
A:Accession: T09223
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-676, 'L', 678-837, 'L', 839-889, 'S', 891-1243, 'D', 1245-1358, 'E', 1360-1368, 'R'
A:Cross-references: EMBL:AF032870; NID:g2760367; PIDN:ACC38820.1; PID:g2760368
C:Genetics:
A:Gene: rhogEF2
A:Cross-references: FlyBase:FBgn0023172
A:Map position: 2; 53Fl-2
A>Note: orchestrates cell shape changes during gastrulation
C/Function:
A:Superfamily: mediates actin rearrangements required for cell shape changes during g
C:Keywords: signal transduction; embryo; GTP exchange
F:1151-1200/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 11.7%; Score 922.5; DB 2; Length 2559;

Best Local Similarity 20.8%; Pred. No. 2.8e-37;

Matches 479; Conservative 279; Mismatches 654; Indels 895; Gaps 82;

QY 17 OSMLSPPIASWLSLSLSDGSPERTSPSHRPSPTSEFTAGLVORC-----VII 67
DB 204 QSNNSNPVLOAPERSKSLNITPLRSLSGHDESTTPATTPSTSLALPKNFQYITLV 263
QY 68 QKDQHFEGFTVSGDRIVLVQSVAPGGAAGVKEGDRIRKNGTWNTSSHLEVVKLIK 127
DB 264 RKDSNNGYGMKVSQDNPNFVSVPGGAETIAGVADMTLRVNGHVRLEKHPVVGLIK 323
QY 128 SGAYVALTLTIGSS---PSPVGV-----SGLOQNPVAVGLVY-----161
DB 324 ASTTVELAVKRSSOKLTPRPSVSVPVTPSPILSGRDTATISGPQPVDSIKRREMETTKIO 383

```

QY 162 -----NP----- 163
Db 384 TLQKMLEKLNLERLKSDNNPXYKLSBANIRKLEOLHOGVADAPTVKLOAAGNKN 443
QY 164 ---LIPP-----PPP----- 170
Db 444 TALLPNOIOHLASATSHNOQFHLHHNNHNNNYPPQOOPASTSPAFLSLPRSLS 503
QY 171 ----- 170
Db 504 LSLGTRKNTKEDLTTSPFGLTDTDLQOORMSHOAESQSMHQTSTPSQFFHPHQ 563
QY 171 -----PPPLP---PPQ-----HTG 182
Db 564 OOHREKETGPTSGKNKFLISRLIEDVPPLPQRNPPROLNLDKNGNASPGSHLYA 623
QY 183 P-----KPLDPEYOKHATOILMM 202
Db 624 PVSDDLDRATSPOLNRSQOOLPRSTNDSPSNKSKRKTIKTALSDP---KASTQFLOM 680
QY 203 ----- 202
Db 681 ESASAGAAGSIEVDGPPPLPPLPGMTEEDMSRSGCONLAQPNVGTAFNYPLVSTT 740
QY 203 -----LROBEELDIL-----PCGERSQR----- 223
Db 741 TAVQNDNLIAEPRLSQRPNIYOOLQYQOQOQOHSOGATGALGOTPLGKKNKRRVGS 800
QY 224 -----TCERLSVDSQEADSGDSTEREPSISELM-----NRNS 259
Db 801 SPDNMHRPRDRITKTTSQSWELVEKDGESS--PRGTPPPLSSHMVLEDPNENRGA 859
QY 260 VUSDGP--LDSPQ-----TSPVILARYAQH--HRRQSDALLPLHOGIDOSPPLII 309
Db 860 AAGPGVFIESHQFTPMAGASSPPISLSHNMHMAOSND-----TOKEIIS 906
QY 310 GREEDD-----PGYNNESDILFODEKLSHPATLVYVRYILSOADGPLLYL 361
Db 907 MEDENSDLDPEPLDENGPPNNLRLLEAE-----NVTPLATPLNIVISNDAPRLFL 960
QY 362 CSEVYOOTNPKDSRLSGKDIWNIFLEKNAPLR--VKIREMLOAEIDLRLNNEP---RN 416
Db 961 ITELVEGTSKDMRKAAVEIHSFVLVPARPLSWYRODESAREVDVULOEDYKVELRT 1020
QY 417 VLCEAOAVMLEIOEOINDYRSKRTIGLSLYGEND--LLGLDGDPLREKMAEKOL--- 471
Db 1021 VELRSKRRAKDLISEQLREFOOKRTAGLTIGYPTDKLAEAKTDLRE--QIIDKYLMPN 1079
QY 472 -----AALGDIISKYEEDRSAPMDFA--VNTFMH----- 499
Db 1080 LHALIDENGSPPEYDKVALCSALSTVYIRIFNTPRPSSIVERVNHVSRKSKSR 1139
QY 500 -----AGIR---LRE-----SRSSCTAETKOS 518
Db 1140 MGNRKMRNVRGHPVLVROYEVTCHNHCOTIIMGVSPQGHCTDKLINIROCSKVDES 1199
QY 519 APDKOKWLPFPPTKKOSSNKKK-----DALDKR----- 551
Db 1200 CGCP-----LPOAKRLAHNDKISKFMGKIRPRTSDVIGNEKRSRODEELNELLPTDRQ 1253
QY 552 -----NPILYIKPKSSOSIKPGANRNIIOHFENSHQYDVEPG----- 592
Db 1254 ASIVRQPSDRRPAANISISNGTSCNTSGL---NTTDLQSSHSGCANDSTIPGGAGC 1310
QY 593 TORLST-----GSFPE---DLLESDDR-----SEIRLGRS 620
Db 1311 NMDLSTSVASTPTSTSGVAAGLSAPAEALNADLTVDKEARRERYSOHPKHSAPVSNMS 1370
QY 621 GSLKGEEMKRSRKAEN-----VRPRSVDVMDAAALAHQASASSASLS----- 668
Db 1371 ESKYERLSNKRNRNRSRKTSDPSLSRP--NDEQLDGLSNATYVGSNSLSAGGTESP 1429
QY 669 TSLKLPPTP-----FTPKMGRRSIESPNL-----GFCITVILPHLEDDLGOLS-- 713

```

```

Db 1430 STSMHEFAAPGAAGVOYPPMGLNONOHPHLLIOHNOQYCCOODSFQAGLAGAASSAAS 1489
QY 714 -----DLPEPEYO---NMOHTYKQDVANULTOREIDREVINELFVT 753
Db 1490 NSSFMNAGHPRLVYARKTLESEDEDDVENDADSSMVAABVSAALTDKAEKKROEINIEYOT 1549
QY 754 EASHLRTKLVLDLIFYOVRMKNEMPREELARLF--PNIPELIEIHSNWCAMKKLR--EEG 811
Db 1550 ERNHVATTKLDRLEFLPLYESGLISODHLLLPALSTLREIGHAFQOSIKORRIEHN 1609
QY 812 PIIRIDSPMLARFQGPRAEELQYVAACRSQVSALELIRKOKRESFOLPMQAESEH 871
Db 1610 HAVNTIGDLAOMFQOSGVNLCERPAQCAQOIALLEALKERKNDEMLOKRLKSES 1669
QY 872 POCSRLOLRDLIVSEMOQRTKYPLELNIITKHT-----EGSTSEHKLRAPDOCREILKF 927
Db 1670 KACRRELKDLPLTYLORTKYPLEFENIKYTVLRLENTEAEALQRAVESKRILVE 1729
QY 928 VNEAVKQTENRHRLSEGYOKRLDATALERASNPLAEFKSLDITTKMHIEGLTRISKD 987
Db 1730 VNQAVRTADAHKLOINQKRLDRSSYDK-----EEFKLDTLOHHLJHDGMLT--IKKN 1781
QY 988 KTLDOVLLEDLVYLLOREERLLK--CHSKTAVOSSDKOTFSPLYKLNA--VLIRSYA 1045
Db 1782 PSVOLHGLLFEMNIVULTYQDDKYIKNHTPLST---TNKPRPIMSIDADTLIROEA 1837
QY 1046 TKRAFFIICITSELGPQIYELVALTSSDKNIMELLEBAVONATKH----- 1092
Db 1838 ADKNSFFLI---KMTSOMLELRAPSSSECKTWFKFHSVDAAROSKRNKSNASHNDYSI 1894
QY 1093 --PGAPIRHPSPGOSQ-----PAYGSTSS 1118
Db 1895 SDPALAIP--HNTYESTLESTDTVOPLAATLTTPLAPMLPIATYTPRATNNNSNVS 1953
QY 1119 -----RVEINDSEVYHNEKPKKLEPGCPGEORVODKOLI--AOGEPYO-----EED 1163
Db 1954 SILTVOLRNPOQDARASEDDAYVNTPRKSSONENRMTSRSGEPLOKXSANGTEAN 2013
QY 1164 EEBELR---TLPRAPSLDGENR-----GIRDRPVLLAL-----TGPLMEG--LA 1204
Db 2014 DVTLRHSQSTRESVPRGSGEERNSTYGVAGNSKRDASIVCSNNSNNTRLMQSPLY 2073
QY 1205 D-ALEDEVNHLHLLMSLPGHTVKTQAAGERD-----DLTPP-----SVYSI 1249
Db 2074 DPTAIOVSTSPAHTEPVLTPGEKLRDLASTIRNDLEKOKITICDIFRLPVHTQOYDI 2133
QY 1250 TSHPMDPGSPGAPRTISDSTRLARPEGSO-----PEGEDVAVSSL---AH- 1291
Db 2134 AMMP-----EAPKDSADIALAAVDOIQTLTKMLNEYHNVPEQGVSAVSTAVCGHCE 2186
QY 1292 -----LPPRTSSGVWDSPELDNRNPALEASTEPRASTYKVRKVSILPPGGVGAAY 1343
Db 2187 KEKLRKKVAPSSSFSSPPRLPPRNQHOAOAOLPPS---RLMPRLQTLDDEN----- 2238
QY 1344 ACSNAIPDSGSESELSEVEGGAQATGNCFYVMPAGPLDSSTETPTGPSPSQCHSLPA 1403
Db 2239 -----AIHEDDDGVCETDEL-----LPAIRKPHRERT--PLAP----- 2272
QY 1404 WPEPOPYRGVGGCCSLVARDVDVIFHTI-----EQLTIKLHRL- 1444
Db 2273 FNTERTSQSV---IDASKRQSTDAVPEGLLEOEPLLEDKTEITGDENEVTPSDKLS 2328
QY 1445 -----KDMELAHLELKSLOGESSGGTTPVGSFHTEAARTDVSLSPPAKALASQONO 1500
Db 2329 ESCNEERQCEVEDITKEVADPTTSKNEAASVDELPSQSRKEIKTEAMKSVADKEDNE 2388
QY 1501 EOGSCPEEGSDIALEDSATDTAVSPGP 1527
Db 2389 E---TIEEGV--ASTVDSSTQTSPTESP 2411

```

RESULT 2

A89633
protein F13B6.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C:Accession: A89633
R:anonymous, The C. elegans Sequencing Consortium.
S:Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: A89633
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-979 <STO>
A:Cross-references: GB:chr_X; PIDN:CAA92119.1; PID:g3875839; GSPDB:GN00028; CESP:F13B6.6
A:Note: similar to Tiam-1 like protein
C:Genetics:
A:Gene: F13B6.6
A:Map position: X

Query Match 8.5%; Score 672; DB 2; Length 979;
Best Local Similarity 27.7%; Pred. No. 1.7e-25;
Matches 223; Conservative 122; Mismatches 270; Indels 190; Gaps 27;
Oy 415 RVVLEAGDAVWLLEIOEIOINDYRSKRTGLGSLYGENDLLGDDGDLREROMAEKQLAAL 474
Db 15 KRIIPFGQRAVTDINMTLNDPFRKKQLG-----GOVGEQ----- 49
Oy 475 GDILSKYEEDRSAPMDVAVNTFMHAGIRLRESRSS-CTAEKTOSAPDKDWLPFPPT- 532
Db 50 --MLFRCEOCNSNSTDF-----DSCETRTQSIISLAVTKISAGA-----LQIHPAS 95
Oy 533 -----KKOSSNKKEKDALEDKRNPIIRYIGKPKSSOSIKGNVNTIIO----- 578
Db 96 LKIVLGTSSSSNEKMLD-----KFPQFARADPPKSKRL--PANKNKPIQYKGHGHSFS 148
Oy 579 -----HFENSHOYDVEPEGTORTLSTGSPFEDLLESRSSE 614
Db 149 INSVATVHYCQCRALINGMQPVIYFCSCNDYKVAHPHOTSALTACYP--VTOSKOKSK 205
Oy 615 IRL-----GRGSLKGREEMKSRKAE-NVPRPSDV-----DMDAA-----EAA 654
Db 206 SRLSLGIRSDANDEDEGSNAHBEQPNVKSTSSDGIGEGHNDRGLVSRSHSMRYAVA 265
Oy 655 RLHQSASSASSLSTRLENTPP-----FTRPK-----MGRRSIES 690
Db 266 TLPOSLSDEKVISPSKRDRAVTPSMORSTGYDLTPADEIDESDIGRKSXYLERRSLES 325
Oy 691 PN-----LGFCTDVLPLHLEDDL-----GOLSDLEPEPEYQNM 724
Db 326 SSRMAIDQSVSAASCHHSGSVAMDDICVRRTTTIOFSRPMGD--SDEIETEARPL 384
Oy 725 QHTYQKDVANLTQREIDROEVINELFVTEASHLFTLRVLDLIFYQMRKKNENLPREELA 784
Db 385 EQLIGMDVIRHLKPE-----KKOLFETERTHYANKILTVHVEFKPIVTSKIIVTEELAN 438
Oy 785 RLPLRLPELIEHNSWCEAMK-----KLRECPRIIRDISPMLARFCGPAREELQ 835
Db 439 LFLANLELLHLHKSMSAMDAEVEKMSAPPRVNGGIYGIYVMESEMFGEAEENLMR 498
Oy 836 VAAOFCYSVALLELIRT--KORRESRFOLPMEASHPOCRRLQRLDLVISEMQRLTKY 893
Db 499 VTATFCQHQHQALEFLRTRCKREKDDAFVRFLAEESNPVCRKIQOLKMIIVEMOQLVKY 558
Oy 894 PLLENTIKHTEGGTSENEKLCRAPDOCRELTKVNEAVKOTENRHRLEGYQKLDATAL 953
Db 559 PLLLETTAKYTTPESEENCLRTVASAKRILISAVNTAKRAENMIRLEELQKRTDSDP 618
Oy 954 ERASNPPLAAEFKSLDTTRKMIHGGPLTWIRISKDKTLDOVLVLLQROEE--RL 1011
Db 619 DKRP--VGHDTNLTNLTFRFLVHDGPDLTCRPNRGKMIELHVVLLLENMVLVLTFTKNSDGNKL 676

Oy 1012 LKCHSKTAVGSSDSKOT-FSPVLKLNALVIRSVATDRAFFICTSELPQIYELVAL 1070
Db 677 VLK-----ALEPSKETRMSPIIPLAPLRIKENDKRAFFLIENSQYQ-AQIYELVAG 728
Oy 1071 TSSDKNIMWELLEBAVONATKHPGA 1095
Db 729 TATERKKTWFKLMGDQIMSEKKNQAA 753
RESULT 3
Rho-guanine nucleotide exchange factor - mouse
T30867
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 02-Sep-2000
C:Accession: T30867
R:Geblik, M.F.B.G.; Poland, M.; Kranenburg, O.; VanHorck, F.P.G.; Moolenaar, W.H.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z20911
A:Accession: T30867
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-1693 <GEB>
A:Cross-references: EMBL:U73199; NID:g1657834; PID:g1657835; PIDN:AAB18197.1
A:Experimental source: Brain
C:Superfamily: protein kinase C zinc-binding repeat homology
F:652-698/Domain: protein kinase C zinc-binding repeat homology <KZM>

Query Match 6.6%; Score 521; DB 2; Length 1693;
Best Local Similarity 20.2%; Pred. No. 1.1e-17;
Matches 290; Conservative 218; Mismatches 496; Indels 430; Gaps 55;
Oy 205 QEEELDILPPCGEFTQRTCEBGLSVDSQPADSGDLSGTERPFSIESLNRKSVSDP 264
Db 283 EQEAKTEKATWPSG-----AAETEVEYNLISG--RSPSEEE--DAKSISQ- 326
Oy 265 GLDSPOTSPIILAVHQAHHROGSDAALLPLNHQIGIOS--PKPLITIGPEEDY----DP 317
Db 327 -VDSP-----SEH-----EDQDRALDRSPFGLLKSKHVPASLAAQGLSDVYLANGDE 372
Oy 318 GYFN-----NESDIIIFDLEKLKSHPAVLYVFLRYLSOADPGPLLYLCEVYQOTN 370
Db 373 VYANCMVIDQVGDDIDVINIEGLSTH-----TS 401
Oy 371 PKDRSLGKDLWNIFLEKNAPLRKYIPBMLQAEITDLRLNNEDPRN-----YLCAOE 423
Db 402 PESGRM-----LGPQACMHTLPDPTSPCGRL--IENSEGTIDAAASQSFYTPSSSR 452
Oy 424 AVMLEIQ-----BOINDYRSKRTGLGSLYGENDLLGDDPL--REROMAEKOLAAL 474
Db 453 TSNLNLSPGLHGFKEKBSHLKRRSSLDALVADSEGGSGSEPLCYAVAGSOSSPRGLPS 512
Oy 475 GDILSKYEEDRSAPMDFAVNTFMS-----H-----AGIRLRESRS-CTAEKTQSPADK 522
Db 513 GDEIDSEFTMTPEPCNISRTESLSLSTLHSESLGIR--SRYSQSSPKIKSGSKR 569
Oy 523 -----DKMLPFPFKKKOSSNKK----- 540
Db 570 LVRFDTVCSYSEBORSYSFQEPPEKRIQEEDEVEYIIPAKSEKRYKVSRTESFLNRM 629
Oy 541 -----KEKALIEDKRN-----PILRYGKPKS--SSQSIKPGVNRNTIIO 579
Db 630 TSPRNKSKMNKDKTEKEKKNRNHQFVPGTSGVYQCGCGCKTILGKESLQACNCK----- 684
Oy 580 FENSHQ--YVPEPGTQRLS--TGSPFEDLLESDSRSRSTIRLGRS-----GSLKG 625
Db 685 -ANTHKCKDAVPPCTCKKFEQYKKNKPOSITLSSSVRDVPAPAGLSHPSSMPIGIPAG 743
Oy 626 REE-----MKRS--RKAENVPRPSDVDMQAAAEALHQA--SSASLSTRSLR 673
Db 744 RKEFAAOVHPLRSVPETTLSEFRRAVTSLESBGDSWRSHSHDELFOQSGSSPSTESF- 802

QY 674 NTPTPTFKMGKRSIESPMLGFCSTVILIRPHLEDDLGQLSDLEPEREVQNMHTGKDV 733

Db 803 -----MMEDVDSIMID-----LSDAQFEAEASWLVVDPSC 837

QY 734 AMLTREIDROEVINELFTTEASHLRTLVLDLIFYQARKENLMPREELAFENLEL 793

Db 838 SHQEKDVAKROVDVIELMQTEVHNITLIMSEVFRKKGKELDQDHSVDKIPCOLDEL 897

QY 794 ITHNSWCEAKMLDEGP-----IIRDISPMLARPDGARELQVAAQFSQSY 846

Db 898 IETHNHFFSMERQESQSGSDRNFTVIMQDILVQDFSEENASKMKRTIYGFPSHKE 957

QY 847 ALELTRKQRKSRFQLFQWAESEHSPQCRRLDRLIYSEMORLTQYPLLENIKHTG 906

Db 958 AMSLKELOQNK-KQNFNIKINSMLNARRGIPROCTILVQRIKPYLVRIQYTE 1016

QY 907 GTSSEHKLCRADQCEILKFNENAVKQTEFNHRELGQYKRLDATALAENAPLAERKS 966

Db 1017 RFEHRDLCKALGILKDMAAVALDKVEYEKQKMLEILNKLENTYTKLKGNHFRQA 1076

QY 967 IDLTRKMHNEGPLTWIRISKDRLDQVLLLEDVLVLRQREKLLKSHKTAAGSSDS 1026

Db 1077 LLSGRALLHDLVYWKATGFRKDIALLMLLDVLLFDEKQKIF-----AANDQ 1128

QY 1027 KQTESPVLKLNANVLRSVATDKRAFFITCTSELGRPOIYELVATLSSDKNIMELLEAV 1086

Db 1129 KPS---VISLOKILIREVANEERGMFLISASAG-PEWEIHTNSKEEENMMRRIOQAV 1184

QY 1087 ONATKHPGAARPIRHPSPGSEPAVQSGTSRVEINSEYVHTEKPKILPEGSPBOR 1146

Db 1185 ESCPREEG-----GRTSE-----SDERKKAERAKI-----QO 1214

QY 1147 VQDKLLIAGFVQDEDEBELTLRPARPSLDENKNGITRPDVLATLGPLMEGLDA 1206

Db 1215 COE-ILSNDDQICTYLEKIL-----IYAEIGEL-----S 1244

QY 1207 ALDEVNLRHLILMSLLRGHTYKKTQAGBERDDLTPTSVVSITGHPMDSGSPQAPTI- 1265

Db 1245 GFEDVHLERHLLI-----KP-----DGEPRDASLIL 1271

QY 1266 -----SDSTRLA-----RREGS-----QREGDVAVSSIANLRPR-TRSS 1299

Db 1272 AALAREASLGVAVASKMGADVQSQSESPGCTVLMDPSTQDVPASPTASLVTEGTGR 1331

QY 1300 GWDSPSELDRNPAAEAASSTEPAASTKYVAKVSLRLPGGVGAAKVAGSNAPISQ----- 1354

Db 1332 GCMDV-----DEGLQGVYTLDAVS-----DAGEVEYR 1359

QY 1355 ---SESELSEVEGEAGATGNCFFVSPAPRLDSS-----TEPTGTPSPSQCHSLPA 1403

Db 1366 STSSGSSQSTIIQALQNLRLYSLQALITQDSHLEINKVLQDQESLAPS-----HSFNG 1415

QY 1404 WPTPEQPYRGVGGCCSLVRDQVDFHTTEQLTFLIKRLKDMELANRELKS 1457

Db 1416 GPLQDQESRYLEKQREELA--NIKHLQHQFOEQRMRHRTCDQOQREQAES 1467

RESULT 4

149342

hypothetical oncogene protein lfc [imported] - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Jun-2000

C:Accession: I49342

R:Whitehead, I.; Kirk, H.; Tognon, C.; Trigo-Gonzalez, G.; Kay, R.

J. Biol. Chem. 270, 18388-18395, 1995

A:Title: Expression cloning of lfc, a novel oncogene with structural similarities to guanine nucleotide-binding proteins

A:Reference number: A57387; MUID:95355462

A:Accession: I49342

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-573 <RES>

A:Cross-references: EMBL:U28495; NID:9902022; PTDN:AAC52234.1; PID:9902023

/Genetics:

A;Gene: lfc
C;Superfamily: pleckstrin repeat homology
F;472-570/Domain: pleckstrin repeat homology <PLK>

Query Match	5.98; Score 462; DB 2; Length 573;
-------------	------------------------------------

Matches 119; Conservative 113; Mismatches 173; Indels 50; Gaps 11;

QY 659 SASSASSASTRSLNPTPPFTPKMGRSIESPNLGECTDVI-----LPHLLEDDL- 709

```

Db 142 SSI SLAKSVSTNIAGHFNDSP-LGLROI LSOS---TDSLNMNRNRTLSVESLIDEGVE 196
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

0V 710 ----G01SDI.EPEP---EVONWOTVGKDVANLT0REIDROEVINELFVTEASHLRTLR 762

```
Db 107 Y E Y N E I M S D E E M D K D E A D S W S I A V D S S E I O O H K K E V M R K R O D V Y E I T O T E I H H V P T I K 256
```

763 WDR TREVODNEVYU WNDDEE ARI EDW DEY TETUNGCGEAWWY DEEC--D-----TT 91A

110

Country	Year	Value	Unit
Algeria	1970	1.0	1000
Algeria	1971	1.0	1000
Algeria	1972	1.0	1000
Algeria	1973	1.0	1000
Algeria	1974	1.0	1000
Algeria	1975	1.0	1000
Algeria	1976	1.0	1000
Algeria	1977	1.0	1000
Algeria	1978	1.0	1000
Algeria	1979	1.0	1000
Algeria	1980	1.0	1000
Algeria	1981	1.0	1000
Algeria	1982	1.0	1000
Algeria	1983	1.0	1000
Algeria	1984	1.0	1000
Algeria	1985	1.0	1000
Algeria	1986	1.0	1000
Algeria	1987	1.0	1000
Algeria	1988	1.0	1000
Algeria	1989	1.0	1000
Algeria	1990	1.0	1000
Algeria	1991	1.0	1000
Algeria	1992	1.0	1000
Algeria	1993	1.0	1000
Algeria	1994	1.0	1000
Algeria	1995	1.0	1000
Algeria	1996	1.0	1000
Algeria	1997	1.0	1000
Algeria	1998	1.0	1000
Algeria	1999	1.0	1000
Algeria	2000	1.0	1000
Algeria	2001	1.0	1000
Algeria	2002	1.0	1000
Algeria	2003	1.0	1000
Algeria	2004	1.0	1000
Algeria	2005	1.0	1000
Algeria	2006	1.0	1000
Algeria	2007	1.0	1000
Algeria	2008	1.0	1000
Algeria	2009	1.0	1000
Algeria	2010	1.0	1000
Algeria	2011	1.0	1000
Algeria	2012	1.0	1000
Algeria	2013	1.0	1000
Algeria	2014	1.0	1000
Algeria	2015	1.0	1000
Algeria	2016	1.0	1000
Algeria	2017	1.0	1000
Algeria	2018	1.0	1000
Algeria	2019	1.0	1000
Algeria	2020	1.0	1000
Algeria	2021	1.0	1000
Algeria	2022	1.0	1000
Algeria	2023	1.0	1000
Algeria	2024	1.0	1000
Algeria	2025	1.0	1000
Algeria	2026	1.0	1000
Algeria	2027	1.0	1000
Algeria	2028	1.0	1000
Algeria	2029	1.0	1000
Algeria	2030	1.0	1000
Algeria	2031	1.0	1000
Algeria	2032	1.0	1000
Algeria	2033	1.0	1000
Algeria	2034	1.0	1000
Algeria	2035	1.0	1000
Algeria	2036	1.0	1000
Algeria	2037	1.0	1000
Algeria	2038	1.0	1000
Algeria	2039	1.0	1000
Algeria	2040	1.0	1000
Algeria	2041	1.0	1000
Algeria	2042	1.0	1000
Algeria	2043	1.0	1000
Algeria	2044	1.0	1000
Algeria	2045	1.0	1000
Algeria	2046	1.0	1000
Algeria	2047	1.0	1000
Algeria	2048	1.0	1000
Algeria	2049	1.0	1000
Algeria	2050	1.0	1000
Algeria	2051	1.0	1000
Algeria	2052	1.0	1000
Algeria	2053	1.0	1000
Algeria	2054	1.0	1000
Algeria	2055	1.0	1000
Algeria	2056	1.0	1000
Algeria	2057	1.0	1000
Algeria	2058	1.0	1000
Algeria	2059	1.0	1000

21 / HANDBUCH DER SOZIOLOGIE UND DER PSYCHOLOGIE 200

8/5 KRQLRDLVSEMQRLTKYPLELLENLIKHTIEGGISEHEKLCRKARDQCREILKFNVAVKQ 934

Db 377 KRHGVQECILLVTQRITKYPVLINRILQNSHGVVEEYQDPLASALGLVKELLSNDQDVHE 436

QY 935 TENRHRLEGYOKRLDATA LERASNP LAE--FKSLDTTRKMIHEGPLTRISKDKTLDL 992

Db 437 LEKEARLQEIYNRM-----PRAQTPVPKGPFGRDELRLKLIHEGCLLWKTATGRFKDV 4922

QY 993 QVLLLEDVLLQREERLLKCHSKTAVGSSDSKQTFSPVLKLNVLIRSVATDKRAF 105

Db 493 LLLMTDVLVFLQEKDQKIYFTSLDKPS-----VSLQNLIVRDIANQAKGMF 540

0Y 1053 ICTSELGPPOIYELVALTSSDKNIMWELLEAVO 1087

Dh 541 TTS---GPFMEYEVHAASPDRTTWIRYTOOSVR 572

RESULT	5
100000	

P47 LBC oncogene - human

```
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
```

C/ACCESSION: 138434
R; Toksoz, D.; Williams, D.A.

A:Title: Novel human oncogene 1bc detected by transfection with distinct h

A:Reference number: I38434; MUID:94119604
A:Accession: I38434

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Residues: 1-424 <RES>

C;Superfamily: CDC24 homology

F:/4-2/1/Donna: CDC24 homology ~CDC24/

Query Match 5.48: Score 428: DB 2: Length 424:

Best Local Similarity 25.78; Pred. No. 5.6e-14;
 Watches 100; Connections 103; Misses 150; Tied 0; Cans 25;

.....

[illegible]

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

[illegible]

Db 98 GVSQGMADLLEQQMVEXLEFCLDELISLHSQFEQRLERKKESLVDKSEKNFLIKRI 137

[illegible]

A:Molecule type: DNA
A:Residues: 1-1158 <MCD>
A:Cross-references: EMBL:AL132828; PIDN:CA860236.1; GSPDB:GN00066; SPDB:SPAC1006.06
A:Experimental source: strain 972h(-); cosmid c106
C:Genetics:
A:Gene: SPDB:SPAC1006.06
A:Map position: 1
A:Introns: 835/1, 975/3
C:Superfamily: CDC24 homology

Query Match	3.3%;	Score 261;	DB 2;	Length 1158;
Best Local Similarity	21.4%;	Pred. No. 4.4e-05;		
Matches 127;	Conservative 106;	Mismatches 279;	Indels 82;	Gaps 16;

QY 461 LERQMAEKQLAALGDILSKYEEDRSAMDPAVNTFMSHAIIRLRSSSOSTAKTQSAP 520
Db 169 ILKHSGSKSASAINIPASHNSNTSSRRPYIHFAFLSQVAVER-----KRLNIC 219
QY 521 DKDKMLPEFPFKTKKQSSSKRKEKDALBKKRNPIARYIGKPKRSSOSIKPGNVNIIOHF 580
Db 220 DAVK-----DGLLYKNAFLGSEAVDLMHIVRTDRRLALGRALDSQMF 266
QY 581 EN---SHQY-DYPEGCTQRLTSGFPEDLLSDSSRSIRLRGSGSLKGREMRSKRAE 636
Db 267 HDVTVSHRLRSLKLEVOYRRIIIPPPLSSMDSGSSI---ENNFLTK---RRANDSD 320
QY 637 NVPRPSVDMDMAAEARLHQSSASSASSLSSTRLE-----NPTPEPTPKGRRSI 688
Db 321 SPDSVLDSOSTPTPISSSVQVNSLAFITSSLSAIRKEAPEDETEVNPBGVFTLLTECYSS 380
QY 689 ESPNLGFCIDVLYLLPLHDDDLGQSLDLP-----EPYQWAGTQVCKDVAN 735
Db 381 TCSRRRLCISISCPRLREQARHLKQDPVLSGSGSTITDQOEDHRLMSENVKQVYDQ 440
QY 736 LTQREIDRQEVINELEVTASHLRTLRVLDFLEYQMKRENILMP---REELAR-LFPNLP 791
Db 441 IDVREMKROEILFEIYTERDFRDLERYIRDFWIKPLSTSAVIPENNROQEIFRCVFHNIM 500
QY 792 ELIELIHNSCEMKRLREGPIIRDISPMU---ARFGPAARELQVAAOFCSQSYAL 848
Db 501 QHHAVNSRLSNALNFTQTLQRPVNTIGDLFDVYKPEEPIKYGANQAIAMF----- 552
QY 849 ELIRTKQKRESRFQLEMOEASHPOCRRLQRLDLIVSEMRQTLTKYPLLENIKHTBEGT 908
Db 553 EFEREKSTNRN-FANVYHEVERLRRESKRLDELNGYLTKEFTRLARYPLLLSGVLKYTKDKN 611
QY 909 SEHEKLCRRDRCRELKFNVAVKOTENRHRHLEBYOKRLDATALERASNPFLAEFKSLD 968
Db 612 PTENIPRTIETRIEFLTKLNTETGKTENRSLILOLNQSLSPADRAK-----LTLF 664
QY 969 LTRKNIHBCPLTWIRSK---DKTLDLQVLLLEDLVYL--LQROEERLLKCH 1016
Db 665 DSRLLIFGVYKVLKASSYSNGDTENDIHMFLLDNFLCLKIKIOMKRRVHKLH 718

RESULT 11
T41524
rhol gdp-gtp exchange protein 1 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C:Accession: T41524
R.Wood, V.; Rajandream, M.A.; Bartell, B.G.; Rieger, M.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z22000
A:Accession: T41524
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1334 <MOO>
A:Cross-references: KMOO;AL049498; PIDN:CA93903.1; GSPDB:GN00068; SPDB:SPCC645.07
A:Experimental source: strain 972h-, cosmid c645
C:Genetics:
A:Gene: SPDB:SPCC645.07


```

Db 483 DEMESTLFGSLPEMLFEOKYFLETL-----BDG--ISAUSDENTLETPSOFRKLLSLGCSF 536
QY 841 CSY-----OSVALELIRTKOREKESRFOLFMQEAESHPCORRLQDRDLIVEMQ 888
Db 537 LYVADHFRLYSGFOANHIKKYQVLEBAKTDKAFKAF--DARNPTKQHSSTLESYLKIPVQ 595
QY 889 RLTYRPLLENIILKHTBEGGSEHEKELCSRARDOCRELKFNENAVAQOTENRRHLEGYOKL 948
Db 596 RFLVYRPLLEKVELSLTDQEESEHYHLTEALKAMEKAVASHINEMOKIYD-----YGVTF 649
QY 949 DATALEASNPDLAEFSLDTLTRKMIHNGCPITW-----RISKDKTLDLOVLLLEDL 1000
Db 650 DQVLAEDS-----GTEKCVLTLSMGELLIMSTYSWLNPELISLCAKAD--LELTIVFKRA 703
QY 1001 VLLLOROEERL--LKCCHSKTAVGSSD---SKOTFSPYLKINAVLIRSVATDKRAFF-- 1052
Db 704 VILVYKCKKLKLLKLPNSRPAHNSSTDLPDFKRWLIPISALOVRIDGPACTENNNSIMEL 763
QY 1053 IICHSLEL--GRPOYLEVALTSSDKNIMMELLEAV--QATNHH-----P 1093
Db 764 IHTKSEIEGRETTFOLCSDSESKNINIVKILSIREFNRHICELPLEKTKCKDRILVP 823
QY 1094 GAAPRIHPSPGSOE-----PAYGSGTSRVEEINDESEVYHTE 1131
Db 824 LKNRPVPSAKIASRSILKVLKUNSSNEMWTGCTGKGLTDSBGSLSSGTOSSGCPTEAGR 883
QY 1132 KEPRKLEGPQPEQRYOD--KOLIAOGEPUOEDEEELTLTRAPRSLDGENRGITRR-- 1187
Db 884 QDSKSTSPGKXPRHGLADFDNLIKESDILSLDEDDDNHRTYVQSGSPTKDITQIOPURLIS 943
QY 1188 -DP 1189
Db 944 EDP 946

```

RESULT 13
T13564
microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*
A:Reference number: Z17689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDD:CAA20006.1
C:Genetics:
A:Cross-references: FLYBase:FBgn0025392
A:Introns: 24/2/; 52/3; 104/3; 179/1; 232/1; 166/3; 256/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: *Drosophila* 576k microtubule-associated protein homolog

[illegible]

Dd	2408	SHANAVS	VTVPTLTKRAELAOIGAKTYS-----	SPLDALRTP----	SAPETISNAD	2457
Oy	184	KPLQ-----	DPEVOKHATQILW-----	NMLROEEELODILPP-----		216
Dd	2458	SPAEKSAEELASOPKSPOVLKESSRPAMVAESKDAAOLKTSVEDLRSPVSTEISRPS				2517
Oy	217	CGETSORFCE-----		GLVSQDGEAD-----	SGLD	241
Dd	2518	AETIASPTEEPADPAEFEDDAEKANULPLTELKMLPLLSBPDYAHASVOPAELSKVD				2577
Oy	242	SGTERFPPISESILMNRSNVSLDPGDSPOTSPTYIARVAOHHRGOSDAALLP-----				294
Dd	2578	IKETASSPIDEA----PKSLIGSPAEEPRP--SPAESAKDAAESVEKSKDASRRPSVEESTK				2633
Oy	295	--LNHQGDIDOSPKLIIIGPEDDDYDGFUNNE-----	SDIIFODLEKLKSHPATLVYFLXKI			348
Dd	2634	ADSTKGDISPSESVLEBPCKDVDEKSESRPPSVASITGSTYKDVSPRASVESVODE				2693
Oy	349	ISOADPGBLTYLCSEVYOQTNPDKSRSLD----	KDWNIIFLEKANAPLVKIPEMLOAI			404
Dd	2694	HJKAKSRESREJAKYESVIDEAGKSDKSSSQDSQDKDEKSYLASKESARESVEESSKDA				2753
Oy	405	DURLNNEDPNRVUCEAO-----	EAVMLEIQOINDRYSKRTL			442
Dd	2754	E---KSESREPESVIASGPVPRESKSPDLSDKDTSRGCVESVTAE--DEKSEQSRES				2808
Oy	443	GLGSUYGE-----	NDLLGIDGDPRLEROM-----	AEKOALAAGDILSKYE		482
Dd	2809	VAESYKAOTKKDGQSQAENSRRSSVDLKKDDDEKORSQOSTIGHKMASTMGD-----				2862
Oy	483	EDRSAKMDFAVNTMSHAGIRLRESRSSCTAECTQAPDKKWLPFPFKTK-----	OSS			537
Dd	2863	---ESPMPKAKDS-----	KEPSRPREVAEIESIKENHTKDEBSPGSKROSVAESIKD			2911
Oy	538	NSKKRKDADLECKKNPLIARYIGRPK-----	SSOSIXKGVGNRNITIQHENSEHOVD			588
Dd	2912	ITKGKSKPLPSKEYVSRPSYVGSIKDKKASRESVAESYKP-----	ESSKDATS			2961
Oy	589	PBPGTORLSTGSFFEDLL-----	ESDSNRSEIRLGRSGSLGREKKRSKAENVPRPS			643
Dd	2962	APPSKEH-----	SPRESYLGSLKDGDKTTSTR-RVSVADSIXIKEKSLIVSQEAS--RPES			3013
Oy	644	DVD--MDAAAEARLHGOSASSASLSSTRBLEMPTPTPPKNGRRTESPNLGFTDYTL				701
Dd	3014	EESISKDKDAAA-----	BQETSREPESYTESVBKGGKSPVASKEASRPAS-----			V 3056
Oy	702	PHILEDDLGOLSDEPEPEVONMOHTYGKVAVANTLTQREIDROEVINELLEVTASHLRTL				761
Dd	3057	ARNAMDSDADESKEQRPESLPQSKAGSI-KDEKSPLASKD-----	EA-----			30966
Oy	762	RVLDLITFORARKENLMRELALFNLPELIELHNSMCBAK-----	KLREGPIIT-RO			816
Dd	3097	-----EKSEKERS--RRESVAQDFPLVKEYVRSPASVAESVKOEAEKSEESPLMKSE				3146
Oy	817	ISDPMLAFRDPAREILOQ-----	VAOFCSQSVALELIR---	TQORKE		858
Dd	3147	ABRP--ASYAASVYDDEAEKSEESRRESVAEKSPDLPSKESRSPASVAESVKODEAKSKREE				3204
Oy	859	SRFOLFQEAESHQOCRRLOLDLIYSEMOR-----	LKYPLLLENITIKH			903
Dd	3205	SRRE---SGAEKSPLASKEASRPASVAESIKDEAEKSEESRRESVAEEKSPPSKEASRP				3261
Oy	904	TGEGIS-----EHKELCRAPOCQE-----	ILNFVANVAYQOTNNRRIRLEGY			944
Dd	3262	TJVAASVYDDEAEKSEESSRDVAEKSPLASKEASRPASVAESVODEAEKSEESRRSV				3321
Oy	945	OKRLDATALERASNPILA-----	AFKSLGIDLTRKMINE-GPLTWIRSKKKTIDLQVL			995
Dd	3322	AKK-SPLAYYKESRNPASVAESINDEXAE-KKESSRRESVAESNPILA--SKASNPSTV-				3375
Oy	996	LLEDVLVILLOBERLLCHKSTANGSSDSKOTFESPVUKLANAVILRSYAIDKRAFITIC				1055
Dd	3376	-AESYKDEAKESKEE-----	SSRQSVAEKSPPLASKEASRPASVAESVODEAEKS			3423


```

OY 1311 PAEEASNEPASYKYVVR-----KVSILPGGVGA--AKVASN-----AIPSGGSESEL 1359
Db 1118 EYMEPPLEESLEAKRVQGLEGRKDLLENAGGIGTFSELPKSKDPMPEPREGHEESA 11777
OY 1360 SEVEGGAO-----TGNCFVSMAPGLDSTE--DTGTPSPSPSOCHSLPAMP--TEP 1408
Db 1178 EAPGAAEAPFAETLGHGGDAPSPMLPGSEAEEDVPVLVSPSPPTPILEDAPGLQP 12373
OY 1409 QPYR-----GVRCGQSSILVRDQVYIFHTTEOL---TIKHLRLKDWELAHRELKSLG 1459
Db 1238 QAEGSQEASMCVQG-----RBAKGVSEODELBSGELPGLOEGEESHEESEDEL 1230
OY 1460 GESSGGTTPVGSF--HTEAARMTDY--SLSPAKE-----ALASQNGQEGGSCPE 15077
Db 1291 GETLPDSTPLGIFYLRPTSRMTPLBESRGHPKLETKETGEGMDPAVLASE---GLEPSEKE 13474
OY 1508 EG-----SDIA--LEDSATDPRAVSPG 1526
Db 1348 EGEGEEBCGKDSJSEFEDLGGTEAPFLDG 1378

```

RESULT 15
JC5795
CDEP protein - human
C.Species: Homo sapiens (man)
C.Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 16-Jul-1999
C.Accession: JC5795
R.Koyano, I.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.
Biochem. Biophys. Res. Commun. 241, 369-375, 1997
A.Title: Molecular cloning and characterization of CDEP, a novel human protein containing
nucleotide binding domain.
A.Reference number: JC5795; MUID:98086358
A.Accession: JC5795
A.Molecule type: mRNA
A.Residues: 1-1045 <KOY>
A.Cross-references: DDBJ:AB008430
C.Comment: This protein is involved in the adhesion, proliferation, and differentiation
C.Superfamily: pleckstrin repeat homology; protein 4.1 membrane-binding domain homology
F1-374/Domain: ezrin-like #status predicted <Ezr>
F442-316/Domain: protein 4.1 membrane-binding domain homology <B41>
F931-1027/Domain: pleckstrin repeat homology <Plk>

Query Match	3.0%;	Score 239.5;	DB 2;	Length 1045;
Best Local Similarity	20.4%;	Pred. No. 0.00043;		
Matches 210;	Conservative 145;	Mismatches 387;	Indels 287;	Gaps 54

0Y	218	GETSRTCEG--RLSVQGEADSGIDSGEREPSTISEJLMN--RNSVLDD--RGDSDPTSP	27 3
Db	2	GEIDORPPPGSRHLCAPENSGISTLERGCKRPPPTSGKLVSTKIOMLDDTOBAFEVFORAP	61
0Y	274	--VILARVAOHNRROGSDAALLPLNHOGIDOSPRLITGDEEDYDPCYFNNESDI----	326
Db	62	GKVLIDANCNH-----LN-----LV---EGDYFGLEFPDHHKITVWLD	96
0Y	327	IFQDIEKIKSHPAVLVFLRYILSQADBRLFLYLCSEYVYQOTNPKQSRSLGKDINWIFL	366
Db	97	LKRPWKQIRRRPKHVY--KEVVK-----FF-----PDPHTLOEBELTRYLF	136
0Y	387	EKNAPLRK-----IREMLQEI--DLRLNNNDPRNVYLCGEQAV	425
Db	137	---ALQYKODLAOGRLTCDNTSALLISHTVQSEIDGFDPEALDREHLANKKYIPODA--	191
0Y	426	MLEIOEONDYRSKRKTGLGSLYGEND-----LLGDDDPRLEROMAEKOLAAL	474
Db	192	---LEDKLVEEHNNH---IGQTPAESDFQLETLRRLEMGIRLHRPKKDEGKINLAYA	245
0Y	475	G-----DILSKYEDRSAPADFAVNTFMSSHAGIRLEERSSC---PAEKTO\$APD-	521
Db	246	NTGILVPGGFLKINAFNNAKYKLSFRKRRL---IKLRPDANSAAQODLLEFLM\$RDF	301
0Y	522	-KDKWL-----PFF-----PKTKQ\$SNSKKEKDALEDKKNRIL-----RYIGK-P	561

```

Db      302  CKSFKICVEHHAFFRLFEERKPKR-----PVLFSGSSFFRSRGRTQ 344
QY      562  KSSOSIKPGVNVNIIOHFENSHQ--YDVEPCTORLSTGSPFDELDSSRSRIILGR 619
Db      345  KOVIDUYEGGHKVV--QEFKRKSHKSHINSLASOPPELNS--EVLBEOOOSTLFTGE 399
QY      620  SGLSLGREEMKRSKRAE-----NVPR--PRSDYMDAAAEAA-----RL 656
Db      400  GAESPCCGSCRGKPEKPVASAGEPCHSPAPRBRSSPAGNKQADGAASAPTEEEEVYKRT 459
QY      657  HQSASSSSASLSTSLNPPPPFPFKKGRSISPNLGCSTVYILPHLLDLOGLDLE 716
Db      460  QOS-----KPPR-----OPSTGSLTG--SPHLSLVSNOGGA 492
QY      717  P-----EPYQNMQHTVGVKDVVYANLQREIDRQE-----YINELFYTE 754
Db      493  PANYLSPNLSPDTKQASPLISPLINDQACPTRDDEGRKRPFDKAVFYAKREVSTTE 552
QY      755  ASHURTLFVLDLITGYOMRKENLMPRELARLPYLPBELIHNMSGEMKRLRE----- 809
Db      553  RYLLKDELVITSMWQSTVSKEDAPREALKLPINPEPLKPFITN--LKLEIQRLLAW 609
QY      810  EG---PIIRD--ISDPMLARFDPAREELQOYVAFCQSYOVALLELITTKORKESRQL 863
Db      610  EGRSMAOIRDYORIGDMWLKIQO---MKHLAAMLMKH--SEALEALENGICKSRRLLEN 663
QY      864  FMOAESHPQOCRLQDLNDLJYEMOORLTQYLLLENITIKHTEGSTSEHKICLRADDORE 923
Db      664  FCRDELOKVC-YLPYLTFLPRLPHRLMNYKQYLERLCKHNHPSNAF-----RD-CRA 715
QY      924  ILKFVNEAVKOTENRH-RLEGYOKRLDATALERASNPAAEFKSLDITRTKMIHNEGPLTW 982
Db      716  ALAETTEVVAOLHGTMIKMFQK-----LHELKLDLIG-IDNLVVPREFIRLGSLSS- 767
QY      983  RISKDKTLTLOVLLLEDLVYILDOERELLKCHSKTAYGASSDSKO-TFSPVLKTLNAVLI 1041
Db      768  KLS-GKGLQORMFELFNDVLLY-----TSKRLTASNOFKYHGQPLXGMTI 812
QY      1042  RSVATDKRAFFIICTSELGPPOIYEIVALTSDKNITMELLEBAVONATKHBGAPRIPIT 1101
Db      813  EE--SEDEMGVPHCLTLRGORQOSTIVASRSRMEKMWEDIOAIDLAKSSSPAPEFLA 870
QY      1102  PSPPSOEP 1110
Db      871  SSPDUNKSP 879

```

Search completed: September 18, 2002, 10:37:04
Job time: 66 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 : Search time 20.02 Seconds
(without alignments)
1863.031 Million cell updates/sec

Title: US-09-695-795-4
Perfect score: 7883
Sequence: 1 MSIRLPHSIDRSKRKQSHL.....EGSDIALSDSATDFAVPCP 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1506.5	19.1	1461	2	US-08-993-228-10
2	1196.5	15.2	912	4	US-08-943-768-2
3	318	4.0	1805	1	US-07-853-913-2
4	258.5	3.3	710	4	US-09-079-812E-2
5	247.5	3.1	1618	1	US-07-853-913-4
6	239	3.0	2101	1	US-08-466-390-4
7	239	3.0	2101	1	US-08-470-850-4
8	239	3.0	2101	1	US-08-467-781-4
9	239	3.0	2101	1	US-08-195-487-4
10	239	3.0	2101	2	US-08-483-924-4
11	239	3.0	2101	5	US-09-452-294-1
12	239	3.0	2101	5	PCT-US93-06160-4
13	237	3.0	626	4	US-09-155-770-7
14	203.5	2.6	1780	1	US-08-769-309A-5
15	203.5	2.6	1780	3	US-08-994-570-5
16	200.5	2.5	1346	2	US-08-635-121-2
17	189.5	2.4	2860	2	US-08-826-267-2
18	188.5	2.4	1104	4	US-08-923-992A-6
19	187.5	2.4	2482	1	US-08-328-254-6
20	187.5	2.4	3248	1	US-08-353-700-1
21	187.5	2.4	3248	5	PCT-US95-16216-1
22	184.5	2.3	1164	4	US-08-923-992A-10
23	184.5	2.3	2466	3	US-09-080-855-12
24	184.5	2.3	2466	5	PCT-US94-09943-2
25	183	2.3	2485	4	US-09-290-640-46
26	179.5	2.3	1886	4	US-08-938-105-3
27	179.5	2.3	2843	1	US-07-741-940-2

28	179.5	2.3	2843	1	US-08-289-548A-2	Sequence 2, Appl
29	179.5	2.3	2843	1	US-08-452-654-2	Sequence 2, Appl
30	179.5	2.3	2843	2	US-08-370-235A-2	Sequence 2, Appl
31	178.5	2.3	1164	4	US-08-923-992A-2	Sequence 2, Appl
32	178	2.3	1939	4	US-09-310-187A-1	Sequence 1, Appl
33	178	2.3	2465	2	US-08-596-291-3	Sequence 3, Appl
34	178	2.3	2465	3	US-09-100-804-3	Sequence 3, Appl
35	177.5	2.3	2842	1	US-07-741-940-7	Sequence 7, Appl
36	177.5	2.3	2842	1	US-08-289-548A-7	Sequence 7, Appl
37	177.5	2.3	2842	1	US-08-452-654-7	Sequence 7, Appl
38	177.5	2.3	2843	1	US-08-452-655B-2	Sequence 2, Appl
39	177.5	2.3	2843	1	US-08-452-655B-7	Sequence 2, Appl
40	177.5	2.3	2843	3	US-08-450-582-2	Sequence 2, Appl
41	177.5	2.3	2843	3	US-08-450-582-7	Sequence 7, Appl
42	177.5	2.3	2973	2	US-08-821-355A-7	Sequence 7, Appl
43	177.5	2.3	2973	2	US-09-003-687A-7	Sequence 7, Appl
44	177.5	2.3	2973	4	US-09-136-605-7	Sequence 7, Appl
45	177	2.2	984	1	US-08-242-932-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-993-228-10
: Sequence 10, Application US/08993228
: Patent No. 5976838
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavallee, Edward R.
: APPLICANT: Racie, Lisa A.
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Agostino, Michael J.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/993,228
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sprunger, Suzanne A.
: REGISTRATION NUMBER: 41,323
: TELEPHONE: (617) 498-8284
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1461 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-993-228-10

Query Match 19.1%; Score 1506.5; DB 2; Length 1461;
Best Local Similarity 29.6%; Pred. No. 3e-105;


```

Db 201 LEMOHTISTDEKSAAYNAIGLYMRHLGVRTKSG-----DKKSGNFRK 247
Qy 530 -----PKTKOSSNKKEDALEDKKRNPIKRTKPKSSOSIKGNVNNIO 578
Db 248 KYMGNNRSDPPKTKKSSSI---LDAARMNGEPQVDPFRHLKAENVAEKKG----- 297
Qy 579 HFENSHQYDVPPEGTQRLSTGSPFEDLLSDSSRSEIRLGRSGSLKGRREKSKAEV 638
Db 298 -----ATDRKGGMV-----SRD-----RNTGAPGQDTPGVSLHPLSL 331
Qy 639 PPRSDVMDAAEAARLHOSASSASSLSTRSLNPPPTPKMGRSIESPNLGFCTD 698
Db 332 DSPDEPGADAPLEL-----GDSPPQPMLESIALPESTDGAET-ESPERG----- 378
Qy 699 VILPHLEDDGQ--LSOLEPER-EVQNMQHTYKDVANLTQRLIDQEVNLEFYEA 755
Db 379 -----DEGEPRSGLEPERPPGMRELVPDTHSLPKSQVKROEVISELLVEA 429
Qy 756 SHLFTLRVLDFIYQMRKENIMPREELARLPNPELIEHNSWCE-AMKKLREGPFI 814
Db 430 AHWRLRLVHLDFPQMAECLEFPLLELONIFPSLDELIVSLDLRMKRRQESGLI 489
Qy 815 RDISPMLARFDPARBELOVAAOFCSTQVALILNTKOKESRFOLPMOAEASHQC 874
Db 490 BEIGVLLARFPGAEGSWFOKISSRFCSKOSFALPOLAKORKDPRCAFVQEAESRPRC 549
Qy 875 RLQRLDLIVSEMRULTYRPLLENIITKTEGTSSEHEKLCARQCEELIKFVNEAVKQ 934
Db 550 RLQQLKDMIPTEMORLTKRPLLOSIGONTTEPT-EREVELAACCBEHLHVNQAVRD 608
Qy 935 TENRRLEGYORKLDATALERASNPAAEFKSLDTTRKMHIEGFLYTRISKDTLDQV 994
Db 609 MEDLRLKDYQRLDLSHROSDDPMLSEFKMLDTKKKLVHEGFLYTRVTDKAVEVHV 668
Qy 995 LILEDVLLQROEERLLKCHSKTAAGSSDSKQTFSPVLKINAVLRSVATDKRAFTI 1054
Db 669 LILDDLILLORODERLLKSHSRFLTPDGTMLRPLRTLSAMTEVATDHAFAFVL 728
Qy 1055 CSEELGPQIYELVALTSSDKINMELLEAVNQ-ATKHGAAPRPHSPGSEPRVQ 1113
Db 729 FTWD-GEAOIYELVAQTVSEKNNMCALITETAGSLKVPAPASPKP-RPRPSTREPLIS 786
Qy 1114 GSTSRVEINDESVYHTEKEPKLPEG-----PGPEORVOD-----KOLIAQSEP 1158
Db 787 SSENQN---GGRETSAPARATERILISDLLPFCRPGEGOLATALRYKLSLKOLFP---P 840
Qy 1159 VOEDEEELRILPRAAPSLIDGENRGITRDVYLLALTGFLMEGLAD-----AALEDVE- 1212
Db 841 AEEDN-----GAGPPPDGD--GVPGGGLSPARTQE-IDENLISLEETWKOLELEE 889
Qy 1213 ---NLRLILMSLLPGHTY 1228
Db 890 EFCRLRPL--SOLGNSY 906

```

RESULT 3
 US-07-853-913-2
 : Sequence 2, Application US/07853913
 : Patent No. 5338839

: GENERAL INFORMATION:
 : APPLICANT: McKay, Ronald D.G.
 : APPLICANT: Lendahl, Urban
 : TITLE OF INVENTION: Nestin Expression As An Indicator of
 : TITLE OF INVENTION: Neuroepithelial Tumors
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
 : STREET: Two Millitia Drive
 : CITY: Lexington
 : STATE: Massachusetts
 : COUNTRY: U.S.A.
 : ZIP: 02173

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/853,913
: FILING DATE: 19920319
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/660,412
: FILING DATE: 22-FEB-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/603,803
: FILING DATE: 25-OCT-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/201,762
: FILING DATE: 02-JUN-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/180,548
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: MIT-4641AAAA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-9540
: TELEFAX: 617-861-6240
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1805 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-853-913-2

```

Query Match 4.0%; Score 318; DB 1; Length 1805;
 Best Local Similarity 20.0%; Pred. No. 5,8e-15;
 Matches 323; Conservative 229; Mismatches 605; Indels 456; Gaps 81;

```

Qy 170 PPPPLPPQHTGPKPLDDPEYQKATQI--LW-NMLROEBEELDILPPCGEFSQKRC- 225
Db 159 PRRPAPPHRIIPGPA-----PEVEDLARLGEVWRGAVDQERYAHMSSSLGQARERLSQ 214
Qy 226 -----EGRLSVDSQEAQSGDSTGERPPSISESLMNNNSVLSQDGLSPQ-TSPVILAR 278
Db 215 AVRGARECRLEVOQLQADR--DSLQERKALEQLRLEGR-----WQRLQATDKFQLAV 265
Qy 279 VAQHRRQSDAALLPLNHQG-----IDQSPKPLIIGPEED--YDPGYFNNESD 325
Db 266 EALQEQKQLOSLQIOLIEGGQQLAHLKMSLSLEVAITYRTILLEAENSRLQTPGR-GSQAS 324
Qy 326 IIFQDLKLSHPALVY---FLAYILSQADP-----GPLFLFCSEVYQOT 369
Db 325 LGFLD-PLK--PNFLGIPEDQYIGSVLPALSPISFSPPLPNTLETPVTAFLKQEFQLA 381
Qy 370 N-----PKDSRSLSKDIWNIFLEKNAP-----LRVKIP--EMLQAEID- 405
Db 362 RPTPLASTPIPPISAPCPPAEVAQAQEVPLSLQTOQAPEPLMLKATVPSSSAIIPLEEE 441
Qy 406 -----LRLNNEDEPRNVLCENQ-----EAVMLEIQDINDY 436
Db 442 PGKQOQHFPDDLTSLATNLNHPHTLPAKDOESSESRYSSIFQDEGOIWEIVKEADI 501
Qy 437 RSKRTLGLSLXGENDLGLDDPLARENQ--MAEQQLAALD--ILS-KYEEDSAPMDF 491
Db 502 EVKVENSSAOKTQES---GLDTEETQDSQGLQKQETTLKALQEBEPLMSLKQNYETAKEN 558
Qy 492 AVNTFMSHAGIRLRSSSCAETKQSAAPDKDKMLPFPKTKKQSSNSK----- 540
Db 559 CNSSTEGHLG-----TLEGPEKEKQIPLKSLBEKNVESKTKLENGVPVUSE 604
Qy 541 ---KEKDALEDK---RNPILRYIGKPKSSOSIKPGNVNIIQ--HFENSHQYDVP 591

```

```

Db 605 LKCEDRTDDELMSPKGLTKRSSLGKESQEVVRPSKGNLMTAFKEESQHPGLGP 664
QY 592 GTORLSTGSPEDLLSDSSRSERILGRSGSLKGREEMKSRKNAVPRRSDVMDMAA 651
Db 665 GAE-----DQMLELVKEDOSFPRSPREEQACRLQKENQPLQEEB----- 711
QY 652 EAARLHQSSASSSLSTRSLNTPPTPKMGRSIESPN--LGF--CTDYLPHLE 706
Db 712 --QOILERLIEKESQESLRS-----PEEQDQAGRSIQKENQEPLEGEEADQOELRLIE 764
QY 707 DDLGQSLDLEPEPEVQWQMTVCQDVANLTQREIDQEVINELFV--TEASHLTRLVLD 765
Db 765 KE--SQESLKSPEE---NORIGKPLERE--NOKSLRYLENOQETFPVLESNORPLKSL 817
QY 766 LIFYQWRKKNLMPRELARLPMLPELIIHNSWCAMKRLPEEGPFIINDISPMLARF 825
Db 818 V-----EEEOQIVKPLEVSO-----DSLQSLAEENV----- 845
QY 826 DGPAREELQOVAQFCYSQVALELITKQKQK-----ESR--FQLEMOAESHPQCRRL 877
Db 846 ----OPLRYLEBDDCINKS-----LLEDKTHKSLGSLERNQDSIIPQSETOVSLRPP 896
QY 878 QLDLDLYSEMQRLLKYLLLENITIKHTEGSTEHEKICLRARDQCRELKLQVNAVQOTEN 937
Db 897 EEBD-----QRTVNHLEKESQEFSSRSSEEOQ-----VMERSLEGEN 933
QY 938 RHRLEGYQK--RLDATALBRASNPDLAEFKSLDTTRKMIHSGPLRWIRISKDTLDQVL 995
Db 934 HESLSVEKEDQWVESQLEKESQDSG--KSLEDSQETF--GPL-----EKENAESIRSL 984
QY 996 LLEDVLLVLOREBERLLKCHSTAVGSSDSKQTFSPVLKLNVLNRSVATDKRAFTIC 1055
Db 985 AGOD-----OEOKEQLEQTOOTLRAVGNQMAVSPPEVDELKPLKLNDE-----I 1032
QY 1056 TSELGPPQIYELVALNLSQDNIM-----MELLEAVQNMTKHGAAP----- 1098
Db 1033 ARSLGKENQSLVSL--KEGCIETVNSLETFIETLETAEDLERKRSIDTOERLMTSEV 1090
QY 1099 -----PIHPSPPQSOEPAYQOSTSRVETINDESVYTHEKEPKKLP--GP--GPEORVQDK 1150
Db 1091 ARETVPEPEDEPPGS-----LGS-----VDENRFTLITSLKESQELSLGKMNVEYTRVEDS 1141
QY 1151 QILAQGP--VOEBDEBELTLPRLAPPSLDGKNR-----GIRKRP 1189
Db 1142 QOOLQVEEGLOEBOHOLEVQOEPLPSSGNQQRWEDVEGKAVQGPAPLATTVGYETDK 1201
QY 1190 VLLALGRLMEGLADALDEVNLRHLI--LMSILBGHVTKQAAGEPDDLTP----- 1242
Db 1202 AELHLKRG---QGGEEBAAEGLLDLDIVGEANSL-----GSSEPKQKVPAPALD 1248
QY 1243 -----TPSVVITSHPMDPGSPQAPPTISDSTR--ARPEGSQPGEDEVAVSSLA 1290
Db 1249 NLBGALLEVVAQSMPEVTERDRDRAQAGQDSIEVTLGLEAARTGLELQEVVAGLEDP 1308
QY 1291 H-----LPRRTSSGCVWSDPELDRNPAEAASTEPAASTKYVRKYSLLPGGVGAAKVA 1344
Db 1309 HFAREEAIRP---SLGE--ESVYKAKIAQGLEGPKEP-----KEAGALDSGLILELPKTS 1357
QY 1345 GSAVAPDSQGSSELSVEGGAQATGKCFVNSPAGPLDSTTEPTGTPPSPO----- 1397
Db 1358 -SEALBCQGHESSE--SMGEWEEBE-----ASLETSDHESGDARQPRPPTTEDEGAQAA 1409
QY 1398 -----CHSLP-----AMPTEDQRYRGVR-----GGQCSLSLRVDVVFHTI 1434
Db 1410 LTAGPKLLEPCSPRIPILTDANHELQQA--EGIDAGQWQPAGSEALERVNEPEFGI 1468
QY 1435 EQGLTLKRLKDMELARRELKSLGESSGCTPRVGSFHTEA--RWT--DYSLSB-- 1486
Db 1469 PE-----GLQDWEEGEESSEADDLGETLPDTPGLYLXSPASPKMDLAGEOQLSPQD 1522
QY 1487 -----PAKEALASDQNGOEGSCPEBSQDIA--LEDSADTDTANSPG 1526

```

```

Db 1523 AGKEDGWPAVPAAGLSGPEEEBEOG-----HGSJDSSEEFEDLGTETASILLPG 1571
RESULT 4
US-09-079-812E-2
; Sequence 2, Application US/09079812E
; Patent No. 6340575
; GENERAL INFORMATION:
; APPLICANT: Bollag, Gideon
; APPLICANT: Crompton, Anne
; APPLICANT: No. 6340575th, Anne
; APPLICANT: Sharma, Sanju
; APPLICANT: Roscoe, William
; TITLE OF INVENTION: Methods and Compositions for Treating Abnormal Cell
; TITLE OF INVENTION: Growth Related to Unwanted Quanine Nucleotide Exchange
; FILE REFERENCE: 1028-US
; CURRENT APPLICATION NUMBER: US/09/079, 812E
; CURRENT FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/049, 879
; PRIOR FILING DATE: 1997-06-17
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO. 2
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Liver Rac GEP
US-09-079-812E-2

Query Match 3.3%; Score 258.5; DB 4; Length 710;
Best Local Similarity 21.3%; Pred. No. 4,3e-11;
Matches 146; Conservative 120; Mismatches 270; Indels 149; Gaps 29;

QY 506 ESSRSCTAEKTQSAPODKW-----LPFPKTKQSSNSKREKDALEDK----- 549
Db 2 ETRESEDELEKTRKKSASDQONTNPEPAKYPELLPEKEESQADQD---IQDEPHCHP 58
QY 550 -KRNPLI-RYI--GKPRSSOSIKPGVNRNIIQHFENSHQYDV----- 589
Db 59 IKRNSIFNRIRKRSKAKANDNERNASCLADSQDGKSVNEPLTINIPWRMPORCTAM 118
QY 590 --EPGQRLS-----TGSFPED---LLESDSSRSEI-----RGRGSGSLK 624
Db 119 QTPDGAQEMKSSSTPGNGATPEEPALADSPYTLTEALRMHP IPADSKRNLIIEQIGLL 178
QY 625 GREEMKSRKAENVPPRSDVMDMAAFAARLHQSSASSSLSTRSLNTPPTPKMG 684
Db 179 YQERYRDKSTQLELTRQODAEIEDNTNGS-----PASQDPEEBEEBEPEEASPE-- 232
QY 685 RSIESPNLGECTDVLPHL---LEDDLQOLSLEPEPEVQNMQHTVQKDVANLTQREI 741
Db 233 KRTL--POLCLLSN---PHGRFNLMODL-----PETRS-----SGVLEIILQPEET 272
QY 742 DREQVINELFWTASHLRLTRVLDLIFYQMRKENLMPRELARLPNLPBELIEHNSMC 801
Db 273 KLOEAMFEIYVTSASYKSLNLLVSHFMENERRIKLILHSEAHILFSNLDVLAVERFL 332
QY 802 EAMKRLREEGPILTRDISDPAARFADGPARELQOVAQFCYSQVALELRLTKQRESRF 861
Db 333 LEIHEHMEENIYVSDVC--IVRY---ADHF--SVITTVYSNQTQYBERITYKOLLOKNAF 387
QY 862 QLEMOAESHPQCRRLQLDLDLYSEMQRLLKYPLLENITIKHTEGSTEHEKICLRARDQ 921
Db 388 RELIAQLELDPPKRCRGLPFSSEFLLIPQRTIRLKLIVQNTIKRYE---ERSERECTALDH 444
QY 922 RE---ILKFYNAVKQTERNHRLEGYQKRLDATALBRASNPDLAEFKSLDTTRKMIHNS 978
Db 445 KELEMYVAKACNEGVKRMKSRTEQMIISQKKME---FKIKSVPIIISHRWLKGQELQOMSG 501
QY 979 PLTWIRISKDKTL--DLOVLLLEDVLLVLR-----QEBRLLIKCHSTAVGSSSDS 1026
Db 502 PKTSRLTKRLKFLHEIYLLFLNDLIVICRQIPDQKQVDSAPRGLLRBE-----LEDO 556

```



```

0Y 1223 LPHGVTKQAG-----BP--EDDLTPPSVITSHPW 1254
      : :||| : ||| :
Db 1026 VEGSAVKGAGLQDPFGSQOQVGTRELAPQGLPEAIPLEYEDVAP-----GC 1075
      : :||| : ||| :
0Y 1255 DPGSP---GQAPITISDTRIAREGSGQPEGEDAVSSLAHLPPRTSSGVWDSBELDRN 1310
      : :||| : ||| :
Db 1076 DQASPEVWLGSPEAMGSAAGAEBCGLQGVG-----GLDDPHLTRE 1117
      : :||| : ||| :
0Y 1311 PAAEASTEPAAKTKYR-----KVSLLREGGYGA--AKVASN---AIPDSGSESEL 1359
      : :||| : ||| :
Db 1118 EYMEPPLEESLEAKRVOGLEGRKDLSEAGGLGTETSELPKSRDPWPPEEGREESBA 1177
      : :||| : ||| :
0Y 1360 SEVEGGAQ-----TGACFVYSPMPAGPLDSTE--PTGTPSPSQCHSPAMP-TEP 1408
      : :||| : ||| :
Db 1178 EAPKGAEEAPFAELTGHTGSDASPMPLSGEAEEDVPPVLYSPSTPIPILEDAPGLQP 1237
      : :||| : ||| :
0Y 1409 QPYR-----GVRCGQSSLVRRDVYIFHTIEQL--TIKLRLKDMELAHRELLKSLG 1459
      : :||| : ||| :
Db 1238 QAEGSQEASMCVQC-----RAAGKVESEQELGSGEIPGLQLEGESESESEDEL 1290
      : :||| : ||| :
0Y 1460 GESSGCTTPVGSF--HTEARMTDY-SLSPAKE-----ALASDSQNOEGSCPE 1507
      : :||| : ||| :
Db 1291 GETLPDSTPIGLFYLRSPTSRPMPLSRGHPLKETGEGMDPAVLASE---GLEPPSEKE 1347
      : :||| : ||| :
0Y 1508 EG-----SDIA--LEDSATDTRAVSPG 1526
      : :||| : ||| :
Db 1348 EGESEGECCGRHSDLSFEFFEDLGTFAEPLPG 1378
      : :||| : ||| :

RESULT 6
US-08-466-390-4
: Sequence 4, Application US/08466390
: Patent No. 568562
:
GENERAL INFORMATION:
:
APPLICANT: TOKRATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,390
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESO, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7100
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

```

Best Locality 19.3%: Pred. No. 7.2e-09:		
Matches 239: Conservative 210: Mismatches 469: Indels 320: Gaps 56:		
QY	246	REPSSISESLMNRNYSVDPGLDSPOTSPIYLARVAO-----HHRGSDAALLP 294
Db	73	KHPSSPECLVSAQKYLEGSELELAKTMILLVHSTWSKSPRDWEQFEKIOAEIAYILK 132
QY	295	---LNRG-----IDQSPKPLIG---PEEDYDGYNNESDIFQDEKLKSHP 338
Db	133	FVLHDHEDGLINLEDLENFLQKAPVSTGCTSPPEELSPSH-QAKREIFLEIQAYASS 191
QY	339	AYLVYFLRYILISOADPGPLLFYCGEVYQOOTNPK---DSRLSGKDIWIFLEKNAFLRY 394
Db	192	SG-----NNFLSGSPASPRGDILOTPQOFQMRILYKOLADERS-NRDELELELAENKLLT 245
QY	395	KLP---EMIQATIDRLRNNEEDPRNVLYCAQAVMLEIQ-BOINDYRSKPLTIGSLG 449
Db	246	EKDAOIAMQOIID-RLA-----LTKKQAASPLEPELELELRKNESLTMRL----- 292
QY	450	ENDLGLDGPRLPRERQMAEKOALAAGDILSKYEEDRSAPMDPAVMTFMSHAGIRLRESRS 509
Db	293	-HETLKQODLTKESQMDRKINOL-----SEENGDSLKRLRFASHIQ-QIQDALN 342
QY	510	SCTAKTQSAIPKDKWLPFPFKTKQSSSKKEKDALEAK---RNPILRTYIGKPKSS 565
Db	343	ELTEHSHKAT---QEWL-----EKQAOLEKELSAALQCKLEENNELLO----- 384
QY	566	OSIKGNVANIIOHENSHQYDVPERG---TORLSTSPFEDDLESDSRSKEITLGRS 620
Db	385	-----GKLSOLEENHLSQIQDNPPOEKGEVLGVDILOETLKQEAATLAANNTOQLQAVENL 439
QY	621	GSLLKREBK-----SKRAENPRRSVDYMDMAAEARLHOS 659
Db	440	ETERRQOEKLLIAERGHFEERKQOULSLITDQSSISINSQAKEELEQASQAHGAULTAQ 499
QY	660	AASSASSISTRSLBNTPPTPTPKMGRSISIESPNLGFCTVDIIPHLLEDLAGLSDP--EP 717
Db	500	VASLSLSELTLN-----ATIQQODDELAG--LKQQAKKQAOQLAQILOQ 542
QY	718	EPEVONMOHTYQKDVVANTTOREIDROEYINELFVTEASHLR-----TLRYDL 766
Db	543	EQASQGLRHQV-EQLSSSLKQKQOOLKEVAEQEATROHQAQOLATPAERERASLIERD- 600
QY	767	IFYQMRKMNIPREIRLAFPNLLELLEIHN-----SMCEAMKYLREGPILIRISDM 821
Db	601	---AALKQDEALEKKEKAAL-ELYQOQLQVAAEARDASQOTSVTOAREAEELSKVEE-- 654
QY	822	LARPDGPAAREELQYAAOFCYSQYVALLE-IRTKQKESRFOLFQOEAESHPOCRRLQLR 880
Db	655	LQACVETARQOEHQAQVAE-----LEQLRSBOOKATERVAQE-----K 697
QY	881	DLIVSEMORLTYPLLENTIYKHTGGGTSSEHKLCRAROCCBEILKFVNAVKOTEN--- 937
Db	698	DOLQOLO-----ALKESLKYVTGSLSEERK--RAALAEBOORCISELAEITRSLVE 748
QY	938	---RRLEGYQRLDATALERASNPLAEFKSLDITTRKIMHGG----- 978
Db	749	QHKRREKLEEREAARKGEGEARLLQDGEHQAEFVLRRELAEVMAAQAHTASECQOLVK 808
QY	979	-PLTYR---ISKDCTLQVLLLEDVYL-----LQROEBRL-LKCHSTAVG 1022
Db	809	EYVAARDGEYDSQOEAEQYGA MFODLMTLKECEKARQOLOEAKVAGISHSLEQTS 868
QY	1023	SSDSQOTSFPYLKLANVILRSY--ATDKAFLITCSLEGPQIYELVATSSDKNIME 1080
Db	869	RQOQR-----LAEILANLARALQOYEKRYARQKLAIDLSTIQ-EKMAATSKF-----VA 917
QY	1081	LLEAVQNAQTKHGAAPIDI--HSPSPGSOEPAV---QG-----STSSHVEINDESVYHT 1130
Db	918	RLETLVRAAGEQOETASRLVKEPARAGDQROEWMLEEOGRQCSQOALQAMEREAQDM 977
QY	1131	EKEPKKL-----PGCPGEQVAVQDKQLI-----AQGPVQE----- 1161


```

1 COUNTRY: USA
2 ZIP: 02109
3
4 COMPUTER READABLE FORM:
5
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
8 OPERATING SYSTEM: PC-DOS/MS-DOS
9 SOFTWARE: Patent In Release #1.0,
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/195,487
12 FILING DATE:
13 CLASSIFICATION: 435
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: US/07/901,701
16 FILING DATE:
17 ATTORNEY/AGENT INFORMATION:
18 NAME: PITCHER ESQ. EDMUND R
19 REGISTRATION NUMBER: 27,829
20 REFERENCE/DOCKET NUMBER: MTP-013
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: 617/248-7000
23 TELEFAX: 617/248-7100
24 INFORMATION FOR SEQ ID NO: 4:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 2101 amino acids
27 TYPE: amino acid
28 TOPOLOGY: linear
29
30 MOLECULE TYPE: protein
31
32 US-08-195-487-4

```

Query Match	3.0%;	Score 239;	DB 1;	Length 2101;
Best Local Similarity	19.3%;	Pred. No. 7.2e-09;		
Matches 239;	Conservative 210;	Mismatches 469;	Indels 320;	Gaps 56;

QY	246	REFPISISELMNRNSVLSDPGIDSTQTSBPVILAVYAO-----HHNRQSDSALLP	294
Db	73	KHPSSPECTVSAQVLEEGSELELAKMTMLLLYHSTMSKSPRDWEQEVYIOAEVLAYLK	1322
QY	295	--LHNG-----IDOSPRLIG---PREDYDPGFNNESITITODELEKLSHP	338
Db	133	FVLDHEGLNLNEDLENFLEKAPVPSTCSSTFPELSPSH-OAKREIRFLELOKVASSS	1911
QY	339	AYLVVFLRYLTISOADPGFLFYLICEVYQOONPK-----DSRSIGKDIWNFLKNAPLRY	394
Db	192	SG-----NNFISGSPASPMGDILOTPQOMRLKKOLADERS-NRDELELELENKKLTT	245
QY	395	KIP---EMLQAEIIDLRLRNNEDEPRNVYICEAOEAVMLEIQ--EOINDRSKRTGLGSLYG	449
Db	246	EKOAOIAMQORID-RLA-----LLNEKQASPLEPKELEBELROKNESLTMRL-----	292
QY	450	ENDLIGLDGDBLRLEROAEKOLAALGDIILSKYEEDRSAPMDFAVNTFMSHAGIRLRESRS	509
Db	293	-HETLKOCODLKEKESQMDRKINQ-----SENGDLSFKLRPASHLO-OLODALN	342
QY	510	SCLEKTKOSAPDKRWLPFPFKTKKQSSNKKEDELEDK-----KNPILRYIGKPKSSS	565
Db	343	ELTEHESKAT--OEWL-----EKOALEKELSAALODKCLEKNEMLIQ-----	384
QY	566	QSIKPGVNRNIIQHFENSHQYDVEPG-----TQRLSTSGFPEDILEDSSRSSEIRLGRS	620
Db	385	-----GKLSOLEEHLISOLDQNPPOEKEVGLDVIQLETLEKQEAATLANNTOLOAREML	439
QY	621	GSLKGREEMK-----RSRKAENVPRPSSDVMDMAAARLHOS	659
Db	440	ETEGQODAKLRLAERGHFEEKEQKSLISLIDGSSITLISQAKEEELBOASQANGARLTAQ	499
QY	660	ASSASSLSLSTRLENPPPTPKMGRRSIESPNIGFCTDVIILPHLEDDLGOLSDT--EP	717
Db	500	VASLTSLETTLN-----ATIQOQDELAG--LKQAKAEKXOALQOTLQOQ	542
QY	718	EPEYQNMQHYGKVAVANLQRELDREBVINELFVTASHLR-----TLRYLD	766
Db	543	BOASQGNRHQV-EOLSSLLKKEQOLKEVLEKQEGATQDHAQOLATAAERASLRERD-6000	

Qy	767	IFBMRKENLMPREELARFLPMLPELIEHN-----SMCEAMKULREBPILINDSPM	821
Db	601	---AALQOLEALEKEKAKI-ETLLOOULOVANBARDSAOTSVTOAKREKALSRVEE--	654
Qy	822	LAREDFAREELOOVAOFCSYOSVALEL-IRTKORESRFOLFMQEAESHPOCRLOLR	880
Db	655	LQACVELTARQOEHQAQVAE-----LELOINSEQOKAIKEREKVAE-----K	697
Qy	881	DLIVSEMRULTKPLLENTIKHTEGTSBHEKICRARDQCRILLKFNVAVKOTEN--	937
Db	698	DOLQEOLO-----AKESILKYTKSLSEBEK-RAADALEQOORCISELKAETRSLYE	748
Qy	938	---BHRLEGYOKRLDATALRASNPILAAEFKSLDLTRKMHG-----	978
Db	749	OKHKEKKELEBEERKGRKGLERLLOJGEAHQATEVYLRRELEEMAAOHTAESCECULYK	808
Qy	979	-PLWVR-----ISKDKTLDLOVILLLEDDLVVL-----LQROEERLL-KCHSKTAVG	1022
Db	809	EVAAMRGYEDSQOGEAQAQAMFQEOULMTLKCECKARQELQEAKEKVAAGIESHSELIS	868
Qy	1023	SSDSKOFFSYVLKINAVLITSV--AIDKRAFILCTSELGPQIYELVALTSSDKNIME	1080
Db	869	ROOKR-----LAEIHAHMLAALQOOVEKEVRAOKLADDLSTLO--EKMAATSKE--VA	917
Qy	1081	LLEEAVONATKRHGCAPITP--HSPSPGSOEPAY-----OG-----STSSRVEINOSEYHT	1130
Db	918	RLETLVKKAGQOQETASRELKVEPAPAGBQRPBMLEQOGRFCFCTQAALQAMEREAQM	977
Qy	1131	EKEPKKL-----PEGGPQRQAVQDKOLI-----AQGEVQOE-----	1161
Db	978	GNELEERLALMESQOGQOQEEERQOEFEVARLTOERKGAQADLALAEKAAAELEWRLONA	1037
Qy	1162	---EDEEELRPLPRAPSPSLDQENGNGITRPVLLATLGPILMGDLADLEVENRHLIL	1219
Db	1038	LNBRQVFATILQERLHAALTEKGS---KOQELAKL-----KGLNPAQIKELBELRQTV-	1087
Qy	1220	WSLLPGHTVKTQAAGEPDDULPTPSVSVTSHPMWDPGSPGAPPTISDSTRLARPEGQOP	1279
Db	1088	---KOLKEOLAKKEKEHAS-----GSGAQ-----SEAGKTEPTGPKL	1122
Qy	1280	EGEDVAVASLALHPRTRSSGVWDSPDLDRNPAEAS	1317
Db	1123	EALRAEVSKLEQOCQKQOE---QAOUSLESLAEAEAS	1156

?
? RESULT 10
? US-08-483-924-4
? Sequence 4, Application US/08483924
? Patent No. 5882876
? GENERAL INFORMATION:
? APPLICANT: TOKRATLY, GARY
? APPLICANT: LIDGARD, GRAHAM P
? TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
? NUMBER OF INVENTIONS: INTERIOR NUCLEAR MATRIX
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: TESTA, HURWITZ & THIBEAULT
? STREET: 125 HIGH STREET
? CITY: BOSTON
? STATE: MA
? COUNTRY: USA
? ZIP: 02110
? COMPUTER READABLE FORM:
? MEDIUM TYPE: FLOPPY disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08-483,924
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:


```

Db      246 EKDAQIAMQORID-RLA-----LINEQAASPLEPEKLELDKDNKESLTMRL----- 292
QY      450 ENDLLGLDDPDLRERMAKQALALGDILSKYEDRSAPMDVAVTFMSHAGIRLRRES 509
Db      293 -HETLKOCODLTKTESQMDKINQL-----SENGDLSFKLEFPASHLO-QLQDALN 342
QY      510 SCTAEKQAPDKKWLPPFPKTKKQSSNKKKEDALEDK-----RNPILRYIGKPKSS 565
Db      343 ELTEHESKAT---QEWL-----EKQAQLEKELSAALQDKCLEKKNELIQ----- 384
QY      566 QSIKGNVNRNIQHFNESHQYVPEPG-----TQRLSTGSFPEDLLESDSSRSSEIRLGRS 620
Db      385 -----GKLSQLEHLSQLQDNPPQEKGEVLGDVQLQETLKQEAATLAANNTOQAARVEML 439
QY      621 GSLKGREEMK-----RSKKAENVPRPSVDVMDAAAEARLHOS 659
Db      440 ETERQOQEKALLERGHFEEREKQOSSLITDLOSSISNLSQAKKELEQAQSHGRRLTAQ 499
QY      660 ASSASASTRSLBNPTPPTPKMGRSIESPNLGFCTDVLPHLELDLQSLDLEP 717
Db      500 VASLISELTLLN-----ATIQODQELAG---LKQAQKQQAQLAQTLQOQ 542
QY      718 BEPVONMHTVKGDVANITQREIDRQEVINELFVTEASHLR-----TLRVLDL 766
Db      543 EQASQGLRHQV-EOLSSSLKQKQOOLKEVAKEQKATRODHAQOLATAAEEREASLRERD- 600
QY      767 IFYQNRKKNLMPREELAFPNLPELLEIHN-----SMCEMKKLREGPIIRISDPM 821
Db      601 ---AAIKQLEALEKEKAAL-ELTQOQLOVANEARDSQTSVTOAKRKAELSRVEE- 654
QY      822 IARFDGPAHEELOQVAAQFCSYQSALEL-ITKQKRESFQOLFQQAESHPOCRLOLR 880
Db      655 IQACVETARQOHEQAQVAF-----LEIQLRSEQQAQKATERVAGQ-----K 697
QY      881 DLIVSEMQRLLTYRPLLEIITHTGEGTSEHKLCRAPQCELLKFNVAENAKQEN--- 937
Db      698 DQLOBOLO-----ALKESLKVTKGSLSEERK--RAADALEEQORCISLKAETRSIVE 748
QY      938 ---RRRLGYQKRLDATATLERSANPLAEFKSLDLTTRKMLHEG----- 978
Db      749 QHKRRKKELEERBARAGKLEALDLOGEAHQAFVLRRELAEMAAOHTAESCEQOLVK 808
QY      979 -PLWNR---ISKDKTLQVLLLELDVYL-----LQROBERLL-LKCHSKTAVG 1022
Db      809 EVAAMRDGVEQOEQAQYGAWFQSOQLMTLKECEKARQELQEAKEKVAQIESHSLQIS 868
QY      1023 SSDSKQTSPIVLKLANVILRSV--ATDKRAFITICTSELGPPQIYELVALTSSDKNIME 1080
Db      869 RQONK-----LAELHANLARALOQVOKEVRAQKLADDLSTLQ--EKMAATSKK---VA 917
QY      1081 LLEEVAONATKHPGAAPLPI--HPSPPGSOEPAY-----OG-----STSRVINEISEVYHT 1130
Db      918 RLETLIVRAAGEQOETASSELVKEPARAGDROPEWLEBOQGROFQSTQALQAMEKEABQM 977
QY      1131 EKEPKKL-----PEGSPGEQVRQDKOLI-----AQGEPPVOE----- 1161
Db      978 GNELERLAALMESQGOQOERGOERREVARLTOERGRAQADLAEKARAELEMLRQNA 1037
QY      1162 --EDDEELRTLPRAPPSLDGENRGIRTRDPVILALITGLPLMEGLADALEDEVENIRHLIT 1219
Db      1038 LNEORVEEATLOEALAHALTEKEG---KDOELAKL-----RGLEAAQIKLELELRQTV- 1087
QY      1220 WSLLPGHVTKQAAGEPBDLTPPVSIVTISHPMWDSPGQAPITISSTLARPEGSGOP 1279
Db      1088 -----KQLEKQALAKKEEHS-----GSGAQ-----SEAAGTEPTGPKL 1122
QY      1280 EGEDVAVSSLAHLPPRTSSSGVWSPEDLRNPAEAAS 1317
Db      1123 EALRAEVSKLEQOCCQOQE-----QADSLERSLEABRAS 1156

```

```

PCT-US93-06160-4
: Sequence 4, Application PC/TUS9306160
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESS: TESTA HURWITZ & THIBEAULT
: STREET: 53 STATE STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/06160
: FILING DATE: 19930621
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESO, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MFP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7100
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US93-06160-4

Query Match      3.0%; Score 239; DB 5; Length 2101;
Best Local Similarity 19.38; Pred. No. 7.2e-09;
Matches 239; Conservative 210; Mismatches 469; Indels 320; Gaps 56;

QY      246 RFPISIESIMRNSVLSPPGIDSPQTSPIILARVAQ-----HHRQSGDALLP 294
Db      73 KHPSPCEIVSAQVILBSSELELAKMTLLYHSTMSKSPROMQEPYKIQALAYILK 132
QY      295 --LNHQG-----IDGSPKPLIIG--PEEDYDPGYFNESDILIFQDLEKLKSH 338
Db      133 FVLHDHEDLNLNEDLNLNLFQAKAPVPSTCSSFPPELSPSH-QAKREIRFLEQKVASS 191
QY      339 AYLAVFLAYILSQADPGPLFLYLCSEVYQQTNP-----DSKSLGKDIWNIFLEKNAPLRY 394
Db      192 SG-----NNFISGSPASMGDILOTPOFQMRRLKQOLADERS-NRDLLELELENRLIT 245
QY      395 KIP---EMLQAEIDLRLRNNEPPRNVLCEQEAVALFIO--EQINDYRSKRTGLGSLYG 449
Db      246 EKDAQIAMQORID-RLA-----LINEQAASPLEPEKLELDKDNKESLTMRL----- 292
QY      450 ENDLLGLDDPDLRERMAKQALALGDILSKYEDRSAPMDFAVTFMSHAGIRLRRES 509
Db      293 -HETLKOCODLTKTESQMDKINQL-----SENGDLSFKLEFPASHLO-QLQDALN 342
QY      510 SCTAEKQAPDKKWLPPFPKTKKQSSNKKKEDALEDK-----RNPILRYIGKPKSS 565
Db      343 ELTEHESKAT---QEWL-----EKQAQLEKELSAALQDKCLEKKNELIQ----- 384
QY      566 QSIKGNVNRNIQHFNESHQYVPEPG-----TQRLSTGSFPEDLLESDSSRSSEIRLGRS 620
Db      385 -----GKLSQLEHLSQLQDNPPQEKGEVLGDVQLQETLKQEAATLAANNTOQAARVEML 439
QY      621 GSLKGREEMK-----RSKKAENVPRPSVDVMDAAAEARLHOS 659

```

```

Db 440 ETERGOEAKTLARHGHEEKEKQOLSSLTIDLOSSISNLSQAKRELQASQAHGARTLAQ 499
QY 660 ASSASSALSTRLENPTPTPKMGKRSIESPNIGFCTDVLPHLLDDJGQLSDL--EP 717
Db 500 VASITSELTITLN-----ATIOQDOELAG--LKOAKKEKAQALQOTLQOQ 542
QY 718 EPEVONNOHVYKGVANLTOREIDROEVINELFVTEASHR-----TLRVLDL 766
Db 543 EQASQGLRHQV-EQLSSLKKEQOOLKEVAKQEQATRODHQOLATAEEREKSLRERD- 600
QY 767 IFYORMKRENIMPREELARLPNLPETIEIHN-----SMCEAMKRLREEGPIINDISDP 821
Db 601 ---AALQLELEKEKAKL-ELIQOOLQVANEARDSAQSVYQAQREKAKLSKVEE-- 654
QY 822 LARFDGAREBLQOVAQFCSYQVSALEL-IRTKQKRESRQFLQMOAESHPQCRILQLR 880
Db 655 LQACVETAREBOHEQAQVAE-----LELOLRSEQOKATEKEVEVAE-----K 697
QY 881 DLIVSEMQRILTKYPLLENIKKHTEGTSHEKLCRARQOCRELKRVNEAVKOTEN--- 937
Db 698 DQLOEOLOQ-----AKESLKVTKGSLKEEKR--RADALLEQOQRCISELKATRSLVE 748
QY 938 ---RHLEGYOKRLDATALERASNPILAEFKSLDITTRKMIHEG-----978
Db 749 QHKEREKELEBERGRGKLEARLLQGEAQAETEVRLRELEAMAAQHTAESCEQOLV 808
QY 979 -PLTW---ISKDKTLDLOVLIEDLVVL-----LQOREBL-LKCHSKTAVG 1022
Db 809 EVAMRDPGYEDSOEEAQQYAMFEQOIMLTKECEKARQEOLEAKKEKAVAGIESHSELOIS 868
QY 1023 SDSKQKFFSVLKLNAVLTISV--ATDKRAFICTSELGPPQIYELVALTSSDKNTWME 1080
Db 869 RQOKK-----LAEHLAMARALQOVQKEVRAOKLADLSTLO--EKMAATSEK--VA 917
QY 1081 LLEAVONATKHBPAPIPIT--HPSPGSQEPAY---QG---STSSRVEINDSEVYHT 1130
Db 918 RLETLVAKAGEQOETASRELIVKEPARAGDQPEWLEBQOQGFQSTQALQAMEREAEQM 977
QY 1131 EKEPKL-----PEGGPGQORVODKOLI-----AGEPQVE-----1161
Db 978 GNELERLRALMESQGOQOBERGQEREVARLTQERGRADLALAEKARAEEMRLQNA 1037
QY 1162 --EDEELRTLPAPPSLDGENRGIRTRDPVLLATGPLMEGLADALDEVENTLRHLIL 1219
Db 1038 LNEGRVETALQERLALALEKES---KQDELAKL-----RGLAQAITELELRQTV- 1087
QY 1220 WSLPGHTVKTQAAGEPEDDLPPTPVYSITSHPMWDPGSPGAQPTISDSTRLAREGSQP 1279
Db 1088 -----KOLKEQLAKKEKEHAS-----GSGAQ-----SEAGRTPEPTGPKL 1122
QY 1280 EGEDVAVSSLAHLPPRTRSSGVWDSPELDNRPAAEAS 1317
Db 1123 EALRAEVSKLQOQCKQOE---QADSLERSLEAERAS 1156

RESULT 13
US-09-155-770-7
; Sequence 7, Application US/09155770A
; Patent No. 6300484
; GENERAL INFORMATION:
; APPLICANT: Duill, David
; TITLE OF INVENTION: DNA ENCODING DP-75 AND A PROCESS FOR ITS USE
; FILE REFERENCE: 200130.418
; CURRENT APPLICATION NUMBER: US/09/155,770A
; CURRENT FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-155-770-7

```

```

Query Match 3.0%; Score 237; DB 4: Length 626;
Best Local Similarity 20.2%; Pred. No. 1,5e-09;
Matches 106; Conservative 106; Mismatches 219; Indels 94; Gaps 18;

QY 740 EIDR-OEVINELFTEASHRITLRLVLDLIFYORMKRNLMPREELARLPNLPETIEIHN 798
Db 21 DADRKRKVIQELVDTEKSYKDSLCELFELVLEPLQNTFTLQDMESLFSLEPMLEFQK 80
QY 799 SMCEAMKRLREEGPIINDISDPMILAREPDGARELOOVAQFCSY-----QSV 846
Db 81 VPLETL-----EDG--ISASSDFMTLETPSOQFRKLFLSLGGSFLYYADHFKLYGFCFANH 134
QY 847 ALELIRTKQKRESRQFLQMOAESHPQCRILQDLIVSEMQRILTKYPLLENIKKTES 906
Db 135 KVQKVLERAKTDKAFKFL--DARNPTQSHSTLESYLIKPVQHVLYKPLKLKELVSLTQD 193
QY 907 GTSHEKLCRARQOCRELKRVNEAVQQTNRRLRLEGQKRLDATALERASNPILAEFKS 966
Db 194 ESEBHNLTLTALAKAMEKVASHINEMOKIYED-----YGVFPOLVAEQS---GTEKEY 243
QY 967 LDITTRKMIHEGPLTW-----RISKDKTLDLOVLIEDLVVLQROBERL--LKCH 1016
Db 244 TELSMGELMHSTVSMNLPFLSLGKARKD--LELTVFVFKRAVILVYKCKKRLKPLPSN 301
QY 1017 SKTAVGSSD-----SKQTFSVLKLNAVLTISVATDKRAF--ITCTSEL--GPPQIYELV 1068
Db 302 SRPAHNSTDLDPEKFRMLIPISALQVRLGNPAGTENNISWELIHTKSELGREGTTFOLC 361
QY 1069 ALTSSDKINIMELLEAV--ONATKH-----PGAAPPIHPSPGSGOE-- 1109
Db 362 CSDESEKTNIVKYIRSLRNFRRHIKELPLEKTCQDLVPLKKNVYPYSAKLASSRLK 421
QY 1110 -----PAYQSTSSRVEINDSEVYHTEKEPKLPEGPGPEQORVOD 1149
Db 422 VLKSSSNEMWTGKTGLDLDSEGLSSGTQSSGCPTAGRODSKSTSPGKYPHGLAD 481
QY 1150 --KQLAAGEPQVEDEELRTLPAPPSLDGENRGIRTR---DP 1189
Db 482 FADNLIKESDILSDEDDHRQTVKQGPSPKMDIEIQFQRLISBDP 526

RESULT 14
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.,
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448
 TELER: 25-3856
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1780 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-769-309A-5

Query Match 2.6%; Score 203.5; DB 1; Length 1780;
 Best Local Similarity 17.0%; Pred. No. 2.7e-06;
 Matches 312; Conservative 259; Mismatches 704; Indels 557; Gaps 73;

29 SLSLSGSLGTPRTSPSHHROPSTSETTAGLVORCVIIQKDQHGFGTSGDRIVLVOS 88
 150 SSESMLBELT-----OPTESQANDIGFKKVFKEV-----GFKFTVKDKTEKEDPT 194
 89 VR-----PGGAKMAGVKEGD-----RIKVGTVTNSHLEYVKL 125
 195 VOLLTVKDEGGAGAGDHOPSLGAGEAASEPKOSTEPKETLKRQSHAEI--- 251
 126 IKSAYAAITLIGSSPPS-----VGVSGLQNPVAGVLRVNPITPPPPPL 174
 252 -----SPPAESQAVEECKEKEKEKESKSA-----ESPT 284
 175 PPHQITGPKPLDPPEVQKHNQ--ILW-----NMLROEEELQDILPPCGETSQRTGGR 228
 285 SPVTSSETG-----STFKKFTQGWAGWKKTSPKRPKEDEVE-----ASEKK 326
 229 LSVDSQSGDLSGSTEREPSSISESLMNRNSVLSDPGLDPSQTSFVILARVAQHHRQGS 288
 327 KQOEKEDVTEDDGKAE-----VASEKLASBGAHQEAEASHBERLAE 372
 289 -DAALLPLNHQ-----GIDOSPRLIIGPEEDYDPGFNNESDIIIFQ----- 329
 373 YEKVELPSEEOVSGSQSEKPAFL--ATEVFDEKIEVQEEVAAVHVSTVEERTEE 429
 330 ---DLEKTL-SHPAYLVFLKRIILSOADPGLLFYLCEVVOQTPKRSRLGKIMNIF 385
 430 OKTEVEETAGSVPAEELVGMADPOEAEPKAEVLKLETCVSGEDPTGADLSPD- 484
 386 LEKNAPLKVKIPEMLOAEIDL-----RLRNNEPRLNCEAOEAWMLEIOEINDYRSKR 440
 485 -EK---VLSPPEGVSEVEMLSQERKRVGQSPKLIKFTSLGKLKLGKKY---KGRK 536
 441 TLGLGSLYENDLGLDGPLREROMAE-----KOLAALGDILSKYEEDRSAPMDF 491
 537 G-GDEESGEHTQVPADPSQEOKGESSASSPEPEITCLEKGLAEVQOD----- 588
 432 AVNTMAGIRLRSSSCSAEKTQSAAPDKDWLPFFPKTKKQSSNKKKEKDALEDKR 551
 589 -----GEAEGATSDEGKKREGVTPMAF--KKMTTPKRVARRPSESDE 631
 552 NPILEYIGPKSSSOSIKPGVNRNIIHFENSHOYDVEPTGORTLSTGSPFEDLESDS 611
 632 DE---LKVKSATLSTESTASTASEMOEEMKGVPEPKPEPKRKADTVTSMEALICVSS 687
 612 RSEIRLGRSGSLKGREEMKRS--RKAENVPRPSDVMDAAEAARLHQSSASSAS- 665
 688 KKRARRRSSDEBEGPRKAMGHDHAKADAGKK-ETGTGILLAGSQEHDPGGSSPQA 746
 666 -----SLST-----KSLENPPTPKMGRSISPMUGCTDVIPLHLEDLG---- 710
 747 GSPTEGEGVSTWSEFKRLVTPRKSSKSLERS-----EDSIAGSGV 788
 711 --QLSDLEPEPEVONMHTVGADVYANLTORERDROE--VINELFVTEASH--LRTLVL 764
 789 EHSITPDTPGAE-ESWV-SIKKFTIGRRKKRDKQKQBDAPVEDGPTGADNEDSDVPAV 846
 765 DLIFQMRKKNLMPRELARLPNLPELIEIHSNCEAMKRLREE----- 810

847 PLSEYDAVERREKMEAQ--AQKGAQPE-----QKATEVSELSSESYHMAAAVADGTR 900
 811 -GPITRIDSDPMLARFQDPAREELQVAAQ--FCSYQVLAELIRTK-----QRK 857
 901 AATIIERSPSMIS--ASVTEPLEQVAEALILEEVLEREVLIAEERPTVTEPLEENR 957
 858 ESRFQLFMOEAESH- -OCRRQLDLVSEMQRLTKPULLEN 899
 958 EARGDTVSEALTPDEAVTAETGAPPLGSEGETEASAEETTEMVASVSLDSDPTTEE 1017
 900 I--IKHTEGTSSEHEKLCRARDQCEILKFVNEAVK-----QTENRHLE 942
 1018 ATPQVEVGGVPIDEQGR---RQGEVLQNAEKVKESQLPRTGGPBDVLQVQRAEAE 1074
 943 GYQKRLDTALERSNPL-----AAEFKLDLTTRKMHEGPLTWIRISKDTLDLQVILE 998
 1075 RPEQAEASGLKKEFDVILKYDAQEAKEPPTQGVQ-----TTPSEFEKAPQVTE 1127
 999 DLVLLQOEERLLLKCHSKTAVGSSDSKQTFSPVLKNAVLRVATDKRAFLICTSE 1058
 1128 SI-----ESSELVITTCQAEFLTAGVKSQEMWEGAIPDSV---ETPTDSE-----TDG 1172
 1059 LGPQIYELVALTSSDKNIMMELLEAVQNTKHPG--AAPIPHP--PPSGQEPAYQGS 1115
 1173 STPVADFAPAGTQKDEIV--EIHEENEVHLVPVGRTEAFAVPAKEREPPAPSSFFVQEE 1230
 1116 TTSREINDSEVYHTEK-----PKKL- 1138
 1231 TKQSKMEDT--LEHTDKSVETVSIISTKTEGTQADQYADEKTKDVPFFEGLEGSIDTG 1289
 1139 -----EG-----PCGEOR---VODKOLIAQCEPV 1159
 1290 ITVSREKTEVALMGEOTEAEACKDDALELQSHAKSPSPSVEREMVQVREKTEAEP 1349
 1160 QEED- -EELRTLPRAAPSIDGENRGIRTRDP 1189
 1350 HVNEKLEHETAVTVSEEVSKQLQTVNVPITIDAKEVSSLEGSPPLGQEEAVCTKIQ 1409
 1190 V-----LLALGPLMELADALADVENLR---HLLMSLPQHTYKTAQAGEPE 1237
 1410 VQSSASFTLPAAEKEEVKILGETANILETGETLEBPAGAHVL-----EKKSSSEKN 1459
 1238 DDLTPTPSVSITSHPMWGPSPQAPRTISDSTRLARPEGSQPEGEADVAVSSLAHLPRTR 1297
 1460 EDPAHNPEDAV---PTGPDQAKSTYPIVATTKKGLSDBEKT----- 1503
 1298 SSGVWDSPELDNRPAEAASTEPAASYKVRKVSILLPGGVGAAKVAGSNAIPD----- 1351
 1504 TSLKWSDEVDQVACQEVKVSVAIE-----DLEPENGILLETGKSSKLWNQNIQTAV 1556
 1352 -----SGQSESELSEFVEGGAATGNCIFYVSMAPRPLDSSTPTPTSPSPSCSLPMPPT 1406
 1557 DQFWRTETATEMLTSELQTA---HVIKADSDAAGQETKEEPEPQASQODEPTTSA 1612
 1407 EPOPYRGVGGCCSILVRVD-----VIFPTIQLTILKRLKMDLAEHLKSGGE 1461
 1613 KEES-ESTAVQASDISDKMSSEAKTMYVEVSGSTYNQQLLEEVLPSE- -EGGG 1667
 1462 SSGGTPVGSFHTEAARWTDVSLSPPAKEALASQNGQEGSC----- 1505
 1668 AGTSKVPEDDDHALAEIRKESLVEPKEDKGDGDVDDPENQNSALADTASGGLTKESPD 1727
 1506 -----PEGSDALED-----SATDPAVSP 1525
 1728 TNGPKQKEKEDAOEVELOEGKRVHSESDKATIP 1759

RESULT 15
 US-08-994-570-5
 ; Sequence 5, Application US/0894570
 ; Patent No. 6090929
 ; GENERAL INFORMATION:

```

: APPLICANT: Scott, John D.
: APPLICANT: Nauert, Brian J.
: APPLICANT: Klauck, Theresa M.
: TITLE OF INVENTION: Protein Binding Domains of Gravin
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower/233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/994,570
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 6090929and, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 27866/33451
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-474-6300
: TELEFAX: 312-474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1780 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-994-570-5

Query Match      2.6%; Score 203.5; DB 3; Length 1780;
Best Local Similarity 17.0%; Pred. No. 2.7e-06;
Matches 312; Conservative 259; Mismatches 704; Indels 557; Gaps 73;

: 29 SLSISGDSPTERTSPSHRPSDSTETAGLVORCVIIQKDHGFGFTVSGDRIVLYOS 88
: 150 SSESNIIELT-----QPTESQANDIGFKVFKV-----GKFTYKAKDKTKPDT 194
: 89 VR-----PGAMAKAGVKEGD--RIIKVNGTMYTNSSHLEVYKL 125
: 195 VOLTYKDEGEGACAGDHPDLSLAGEAASKSEPKOSTEKREPTLKRQSHAEI--- 251
: 126 IKSQATVAILTLGSSPPS-----GVVSGLOQNSVAGVLRVNPPIPPPPPPPL 174
: 252 -----SPPASGQAVECKEKEGKEKPEPSKA-----ESPT 284
: 175 PRPHITGPKPLDPPVQKHATQ--ILM-----NMLRQEEELDILIRPCGETSORTCEGR 228
: 285 SPVTSSTG-----STFKKFTTQGAAGMKKTSFKRKPEDEV-----ASEKK 326
: 229 LSVDSQADSGLDSTERTSPISSELMNNSVLSDPQSPQTSPIVLARVAQHRRQGS 288
: 327 KRQEPKVDTEEDGKAE-----VASEKLTGASEQAHPOEPASAEHPRLSAE 372
: 289 -DAALLPLNHQ-----GIDQSPKPLITGPEEDYDPCYNNESDITFQ----- 329
: 373 YEKVELPSEQVSGSGPSEEPAPL--ATEVFDEKIEVHOEEVVAEYHVSVEERTEE 429
: 330 ---DLEKLR-SHPAYLVFLRYILTSQADPGPLFYLCSEVYOOTNPKDSRLSGKDIWNIF 385
: 430 QATEVBETAGSVPAEELVGMDAEPQAPAKELVKLETCVSGEDPTQAGADSPD----- 484
: 386 LKKNAPLRVYIKIMLOAEIDL-----RLRNEDPRNVILCEAOEAVVLEIOEINDYRSKR 440
: 485 -EK-----VLSPKPEGVYSEVEMLSQERMKVQSGPLKKLFTSTGLKLSKKQ-----KGRK 536

```

```

QY 441 TLGAGSLYGENDLGLDGPPLRERQMAE-----KQIALAGDILSYEDRSAPMDF 491
: 537 G-GGDESGEHHTQVPADSPDSQDEQKGESASAPPEEPETLEKELAEVQD----- 588
QY 492 AVNTFMSHAGIRLRERSSTAKTOSAPDKKWLFFPFKTKQSSNSKKEKDAEDKKR 551
: 589 -----GAEEGATSDGKKRREGVTPMAF-----KKMVPKRVRRPSSDKR 631
QY 552 NPILRYIGKPKSSOSIKKGNVNIIOHFNESHQYVVPPEGTORLSTGSFPEDILSDSS 611
: 632 DE-----LDKYSKATLSTESTASSEMGEKKSVEEPKPEEPKRVTVSWELLICVSS 687
QY 612 KSEIRLGRSGSLKGREEMKRS--RKAENVPRPSDVMDMAAARLHOSASSAS----- 665
: 688 KKRARRRRSSDEEGCPKAGDGHQKADENAGKDK-ETGTGDIILAGSOEHPGCGSSSPEDQ 746
QY 666 -----SIST-----KSLNPPPTFPKNGRRIESPNGICFDVILLPHLEDDLG---- 710
: 747 GSPTGEGSVTWSSEFKRLVTPRKSKSKELENS-----EDSINGSV 788
QY 711 --QLSDLEPEPEYQNMQHTVGKDVVANLTQREIDROE--VINELFYTEASH--LRTLYL 764
: 789 EHSTPTEPEGKE-ESNV-SIKKFIPIRRKKRPDQKQEQAPVEDAGPTGANNEDSDVPAYV 846
QY 765 DLIFTQMRKKNIMPBEELARLFPNLPELIEIHSNCEAMKRIREE----- 810
: 847 PLSEYDAVEREKMEAOQ--AKQGAQPE-----OKAATVSKSELSQVHMAAAVADCTR 900
QY 811 -GPILIDISDPMLARFDGFARELOQVAAO--FCSYQSALEELIRK-----QKR 857
: 901 AATIEERSPSWIS--ASVTEPLEVEAEALITIEVIEREYIAEEPPYTEPIPEKR 957
QY 858 ESRFOLFMOEASHP-----QCRRLDLRLIVSEMORLTKYPLLEN 899
: 958 EARGDTVYSEAEELTPBAVTAETAGPLGSEEGTEASAABETTEMVAVSGLDSDPTTEE 1017
QY 900 I--IKHTEGTSHEHKLCAKADQCREILKFVNEAVR-----QENRHRLE 942
: 1018 APVPOVEEGVPDIEQER--RTQEVLOAVAEKVEESQLPRTGTPEDVLPVQVAAE 1074
QY 943 GYKRLDATALERASNP-----AAEFKSLDLTRKRIHSGPLTWRSKKTLDLQVILLE 998
: 1075 RPEQQAFAAGIKKEDVYLVKVAQAEKTERPFOGKAVGQ-----TTSESEKAPQVTE 1127
QY 999 DLVYLLQROEBRLILKCHSKTAVGSSDKOTSPVLKLNLAULIRSVATDKRAFFITCSE 1058
: 1128 SI-----ESSELVYTCQAETLAGVKSQEMVMEQAIIPDSV--ETPTDSE-----TDG 1172
QY 1059 LCPQIYELVALTSSDKNTIMMELLEBAVONATKHNQ--AAPPIHRS--PRSGQEPAYQGS 1115
: 1173 STPVADFDAPGTQKDEIV--ETIHEENVHLVPRGTAEAVPAOKERPPAPSSFFVQEE 1230
QY 1116 TSSRVLEINSEVYHTEKE-----PKKLP----- 1138
: 1231 TKEQSKMEDT-LEHTDKESVETVSTLSTKEGTQEAADQYADEKTKOVPPFEGLSIDIQ 1289
QY 1139 -----EG-----PGPDR--VODKOLLAQGEV 1159
: 1290 ITVSRKETEVALKGGTEAECKKDDALELOSHAKSPSPYPEREVMVQVVEREKTEAEPT 1349
QY 1160 QEEDF-----ELNLTLPRAPSLDGENKRTITRNP 1189
: 1350 HNEEKLHEHTAVTVSEVSKOLLQTVNPIIDGAKEVSSLEGSPPCLGQEDAAVCTKIQ 1409
QY 1190 V-----LALTGPLIMEGLADALAEVENLR-----HLIIMSLPGHTVTOAGAEP 1237
: 1410 VOSSEASFLLTAAEBEKVLGETANILEGTLEPRGALIV-----EKSSEKN 1459
QY 1238 DDLTPPSVVSITSHMPDGPBGQAFTISDSTRLARPEGSQDEGDVAVSSLAHLPPRTR 1297
: 1460 EDFAAHPGEDAV--PTGDCQAKSTPVIVSATTKKGLSSDLEGEKT----- 1503

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 ; Search time 45.49 Seconds
(without alignments)
3728.505 Million cell updates/sec

Title: US-09-695-795-4
Perfect score: 7883
Sequence: 1 MSIRLPHSIDRSASKQSHL.....ESDIALEDSARDTAVSPGP 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7883	100.0	1527	22	AAU01184
2	6455.5	81.9	1520	20	AAV41010
3	1506.5	19.1	1461	19	AAW64468
4	1506.5	19.1	1461	22	AAW64468
5	1216.5	15.4	919	20	AAW41011
6	1196.5	15.2	912	19	AAW42103
7	1194.5	15.2	912	20	AAV41009
8	1100.5	14.0	846	22	AAW12310
9	1100.5	14.0	846	22	AAW79614
10	1100.5	14.0	846	22	AAW41561
11	1068	13.5	848	22	ABG18434

12	945.5	12.0	2559	20	AAV41012	Amino acid sequenc
13	945	12.0	203	22	AAU21680	Novel human neopla
14	936.5	11.9	2559	22	ABW62745	Drosophila melanog
15	892.5	11.3	696	22	AAW39775	Human polypeptide
16	891.5	11.3	696	22	AAW78630	Human polypeptide
17	700.5	8.9	181	22	AAU17436	Human signal trans
18	700.5	8.9	181	22	AAU21805	Novel human neopla
19	541	6.9	123	20	AAV41007	Human KIAA380 prot
20	539.5	6.8	169	22	AAW1141	Human immune/haema
21	531	6.7	1050	22	ABW44551	Human wound healin
22	517.5	6.6	1429	20	AAW3941	Human btx protein.
23	517.5	6.6	1429	22	ABG05357	Novel human diagno
24	513.5	6.5	1066	22	ABG15870	Novel human diagno
25	504	6.4	963	22	AAU17312	Novel signal trans
26	433	5.5	96	20	AAV41016	Human KIAA0380 pro
27	428	5.4	418	21	AAW58227	Lung cancer associ
28	410.5	5.2	339	22	AAO10352	Human polypeptide
29	370	4.7	936	22	ABG20576	Human human diagno
30	346	4.4	1309	22	ABW17862	Drosophila melanog
31	339.5	4.3	1715	21	AAV57449	Mouse Eesl protein
32	332	4.2	235	22	AAW25470	Human protein sequ
33	328	4.2	1697	22	AAW79199	Human protein sequ
34	322.5	4.1	1683	21	AAV71160	Rat phosphodiester
35	318	4.0	1805	15	AAW71204	Rat nestin. Ratu
36	318	4.0	1805	15	AAW60126	Rat nestin protein
37	310.5	3.9	790	22	AAW3045	Human protein sequ
38	301.5	3.8	1658	21	AAV57450	Mouse Eesl protei
39	293.5	3.7	1234	22	ABW68510	Drosophila melanog
40	290	3.7	2189	22	AAW79943	Human protein sequ
41	288	3.7	2161	22	AAW78959	Human protein sequ
42	288	3.7	2523	22	AAU03503	Human protein kina
43	280.5	3.6	558	20	AAW73379	CSB2 protein sequ
44	279.5	3.5	522	22	AAW95513	Human protein sequ
45	274.5	3.5	1576	22	ABG06283	Novel human diagno

ALIGNMENTS

RESULT 1	AAU01184	standard; Protein: 1527 AA.
XX	AAU01184	
AC	AAU01184	
XX	AAU01184	
DT	07-SEP-2001	(first entry)
XX		
DE	Rat glutamate transporter associated protein GTRAP4-48.	
KW	Rat: glutamate transporter associated protein; GTRAP4-48;	
KW	glutamate transporter protein; chloride transport; neurodegeneration;	
KW	cytoskeletal stability; nervous system disorder; schizophrenia;	
KW	spinocerebellar ataxia type I; SCAT; GABA metabolism.	
OS	Rattus sp.	
XX		
PN	W0200130968-A2.	
XX		
PD	03-MAY-2001.	
XX		
PF	23-OCT-2000; 2000WO-US29431.	
XX		
PR	23-OCT-1999; 99US-0161007.	
PR	22-MAY-2000; 2000US-0206157.	
PA	(UYU) UNIV JOHNS HOPKINS SCHOOI MEDICINE.	
PI	Rothstein JD, Jackson M, Lin G, Law R, Orlov I;	
XX		
XX	WPI: 2001-300498/31.	
DR	N-PSDB: AAS04274.	
XX		
PT	Novel substantially pure glutamate transporter associated polypeptide	

PT which modulates intracellular glutamate transport, interacts with
PT glutamate transporter protein and has expression pattern in brain
PS Claim 8; Fig 17; 116pp; English.
XX
CC The present sequence representing novel rat glutamate transporter
CC associated protein GTRAP-48 modulates intracellular glutamate
CC transport, interacts with a glutamate transporter protein and has an
CC expression pattern in the brain. The GTRAP proteins are also involved
CC in mediating chloride transport and cytoskeletal stability. The
CC invention also provides methods for identifying a compound that
CC modulates a cellular response mediated by a GTRAP protein and a
CC compound that inhibits the interaction of a glutamate transporter
CC protein and a GTRAP protein. A compound which modulates the activity
CC of a glutamate transporter associated polypeptide or interaction with
CC a glutamate transporter protein is useful for treating a disorder
CC associated with glutamate transport or chloride flux, preferably a
CC nervous system disorder such as neurodegeneration, spinocerebellar
CC ataxia type 1 (SCA1), schizophrenia, epilepsy or a disorder of GABA
CC metabolism. Such compound could be an antisense oligonucleotide
CC (AA504278) which can be used for modulating glutamate transport in a
CC subject.
CC
XX
SQ Sequence 1527 AA:

Query Match 100.0%; Score 7883; DB 22; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLRHSIDRSASKROSHISPIASWLSLSISLSDTPERTSPSHHROPSDSETTAGL 60
DB 1 MSRLRHSIDRSASKROSHISPIASWLSLSISLSDTPERTSPSHHROPSDSETTAGL 60
QY 61 VQRCAVLIQKQDFGFTVSGDRIVLVQSVRPGAGAMKAGYKESDRITIKVGTWNTSNHL 120
DB 61 VQRCAVLIQKQDFGFTVSGDRIVLVQSVRPGAGAMKAGYKESDRITIKVGTWNTSNHL 120
QY 121 EYVKLIKSGAVYALTLTGSSPSVGVSGIQANPSVAGVLRVNPITPPPPPPPLPPPPQHT 180
DB 121 EYVKLIKSGAVYALTLTGSSPSVGVSGIQANPSVAGVLRVNPITPPPPPPPLPPPPQHT 180
QY 121 EVYKLIKSGAVYALTLTGSSPSVGVSGIQANPSVAGVLRVNPITPPPPPPPLPPPPQHT 180
DB 121 EVYKLIKSGAVYALTLTGSSPSVGVSGIQANPSVAGVLRVNPITPPPPPPPLPPPPQHT 180
QY 181 TGKPLPDPEVOKHATQILNMNMLROEERELDITLPCGERTSOTRCGRSLVDSQEDSGL 240
DB 181 TGKPLPDPEVOKHATQILNMNMLROEERELDITLPCGERTSOTRCGRSLVDSQEDSGL 240
QY 181 TGKPLPDPEVOKHATQILNMNMLROEERELDITLPCGERTSOTRCGRSLVDSQEDSGL 240
DB 181 TGKPLPDPEVOKHATQILNMNMLROEERELDITLPCGERTSOTRCGRSLVDSQEDSGL 240
QY 241 DSGTERPSSISELMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300
DB 241 DSGTERPSSISELMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300
QY 301 DSGTERPSSISELMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300
DB 301 DSGTERPSSISELMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300
QY 301 DSGTERPSSISELMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300
DB 301 DSGTERPSSISELMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300
QY 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLEVKIPEMIAQEIIDLRLNRNEPRNVLC 420
DB 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLEVKIPEMIAQEIIDLRLNRNEPRNVLC 420
QY 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLEVKIPEMIAQEIIDLRLNRNEPRNVLC 420
DB 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLEVKIPEMIAQEIIDLRLNRNEPRNVLC 420
QY 421 AGEAVMLEIOEIOINDYRSKRTGLGSLYGENDLLGLDGPRLRERQNAEKOLAALGDTLSK 480
DB 421 AGEAVMLEIOEIOINDYRSKRTGLGSLYGENDLLGLDGPRLRERQNAEKOLAALGDTLSK 480
QY 481 YEEDRRAPMDFAVNTFMNHAAGIRLRSRSCTAEKQASADKKKWLPRFFTKKQSSNSK 540
DB 481 YEEDRRAPMDFAVNTFMNHAAGIRLRSRSCTAEKQASADKKKWLPRFFTKKQSSNSK 540
QY 541 KEKDALIEDKKRNPIILRYIGKPKSSSOSIKPGANRNIIQHENSQOVDPEPTGQRLSTGS 600
DB 541 KEKDALIEDKKRNPIILRYIGKPKSSSOSIKPGANRNIIQHENSQOVDPEPTGQRLSTGS 600
QY 601 FPEDLLSOSRSRIRLGRSGSLKGREEMKSRKAENVPRPSDVMDAAEARLHQSA 660
DB 601 FPEDLLSOSRSRIRLGRSGSLKGREEMKSRKAENVPRPSDVMDAAEARLHQSA 660

QY 661 SSSASSLSTFSLLENPPPTPKMGRSISPSNLGCTDVLPLHLLEDJGQSLDEPEPE 720
DB 661 SSSASSLSTFSLLENPPPTPKMGRSISPSNLGCTDVLPLHLLEDJGQSLDEPEPE 720
QY 721 VQWQHTVQKDVVAVANTQREIDROEVINELFYEAASHLRTLRDLTLFQRMKKNLMR 780
DB 721 VQWQHTVQKDVVAVANTQREIDROEVINELFYEAASHLRTLRDLTLFQRMKKNLMR 780
QY 781 EELARFPNLPBELIFHNWCAMKRLREGEPIIRDISPMILARFGPAREELOQVAAOF 840
DB 781 EELARFPNLPBELIFHNWCAMKRLREGEPIIRDISPMILARFGPAREELOQVAAOF 840
QY 841 CSYQVALELIRTKOKESRFQLEWQASHPCQRLQLRDLIVSMQRITKYPILLENI 900
DB 841 CSYQVALELIRTKOKESRFQLEWQASHPCQRLQLRDLIVSMQRITKYPILLENI 900
QY 901 IKHTEGTSHEHELCARPOCRRLKFNVAQVQNTNRRLRERQKRLATLERSNPL 960
DB 901 IKHTEGTSHEHELCARPOCRRLKFNVAQVQNTNRRLRERQKRLATLERSNPL 960
QY 961 AAEFKSLDLITTRKMIHEGPIITWRISKDKTLDLOVLLLEDLVVLQROEERLLKCHSKTA 1020
DB 961 AAEFKSLDLITTRKMIHEGPIITWRISKDKTLDLOVLLLEDLVVLQROEERLLKCHSKTA 1020
QY 1021 VGSDDSKOTFFPYVLKLNANVLINSVATDKRAFTICTSELGPPQIYELVALTSSDKITME 1080
DB 1021 VGSDDSKOTFFPYVLKLNANVLINSVATDKRAFTICTSELGPPQIYELVALTSSDKITME 1080
QY 1081 LLEAVQNTKHPGAPRIPHPSPGSOEPAYQGSRSRREINDSEVYTERKPKKLPG 1140
DB 1081 LLEAVQNTKHPGAPRIPHPSPGSOEPAYQGSRSRREINDSEVYTERKPKKLPG 1140
QY 1141 PGEPQVQDKQILAQGEFVQEEDEBELRLPRAPLSDGENRGIRTRDPVLLATGPLLM 1200
DB 1141 PGEPQVQDKQILAQGEFVQEEDEBELRLPRAPLSDGENRGIRTRDPVLLATGPLLM 1200
QY 1201 EGLADALDEVNLRHLILMSLLRGHTVKTQANGPEBDLTPRPSVYSTSHPMWDSGP 1260
DB 1201 EGLADALDEVNLRHLILMSLLRGHTVKTQANGPEBDLTPRPSVYSTSHPMWDSGP 1260
QY 1261 QAPTISDSTRLARPESSQEPEDVAVSSLAHLPPRTSSGVWDSPELDNRPAALASTEP 1320
DB 1261 QAPTISDSTRLARPESSQEPEDVAVSSLAHLPPRTSSGVWDSPELDNRPAALASTEP 1320
QY 1321 AASYKVRKVSLLPGGCVGAAYAGSNALPDSQSSSELSVEVGGQAQATGNCPTYSMPAG 1380
DB 1321 AASYKVRKVSLLPGGCVGAAYAGSNALPDSQSSSELSVEVGGQAQATGNCPTYSMPAG 1380
QY 1381 PLDSSTPEPTGTPSPSOSCHSLPMPTEPQYRGVRCGCSLSVRRVDVIFPTIQLTLTK 1440
DB 1381 PLDSSTPEPTGTPSPSOSCHSLPMPTEPQYRGVRCGCSLSVRRVDVIFPTIQLTLTK 1440
QY 1441 LHRLEKMEIAHRELKSLSGESSGCTTPVGSFTEAARTDYSLSPPAKAALASQONCO 1500
DB 1441 LHRLEKMEIAHRELKSLSGESSGCTTPVGSFTEAARTDYSLSPPAKAALASQONCO 1500
QY 1501 EGSCPEEGSDIALEDSATDTAVSPGP 1527
DB 1501 EGSCPEEGSDIALEDSATDTAVSPGP 1527

RESULT 2
AA41010
ID AA41010 standard; Protein; 1520 AA.
XX
AC AA41010;
XX
DT 06-DEC-1999 (first entry)
XX
DE Amino acid sequence of KIAA380.
XX
KW RGS; guanine nucleotide exchange factor; GEF; RGS domain; GAS; cancer;

KW G protein alpha subunit; cell proliferation; growth control; hemostasis;
KW morphogenesis; stress fiber formation; integrin-mediated interaction;
KW embryonic development; tumor cell growth; cell death; leukocyte homing;
KW bone resorption; clot retraction; db1 homology domain; mechanical stress;
KW pleckstrin homology domain; KIAA380.

XX Homo sapiens.

PN MO9947557-A2.

XX 23-SEP-1999.

XX 18-MAR-1999: 99WO-US06051.

XX 18-MAR-1998: 98US-0078634.

XX (ONYX-) ONYX PHARM INC.

PI Bollag G, Hart MJ, Roscoe W, Polakis P, Sternweis P, Kozasa T;
PI Jiang X;

XX WPI: 1999-571821/48.

DR N-PSDB: AAZ22282.

XX New isolated RGS-GEF polypeptides, used to develop products for
PT modulating, e.g. cell proliferation and integrin-mediated interactions

PS Examples: Fig 12: 75pp; English.

XX The invention relates to isolated RGS-guanine nucleotide exchange factor
CC (GEF) polypeptides. The novel isolated RGS-GEF polypeptide consists of an
CC RGS domain of a GEF protein and does not comprise a db1 homology (DH)
CC domain or a pleckstrin homology (PH) domain. The RGS-GEF polypeptides can
CC be used for modulating an activity of a G protein alpha subunit (GAS).
CC The products can be used for the regulation of biological pathways in
CC which a RGS-GEF polypeptide is involved, particularly pathological
CC conditions, e.g. cell proliferation (e.g. cancer), morphogenesis, growth
CC control, stress fiber formation, and integrin-mediated interactions, such
CC as embryonic development, tumor cell growth and metastasis, programmed
CC cell death, hemostasis, leukocyte homing and activation, bone resorption,
CC clot retraction, and the response of cells to mechanical stress. The
CC products can also be used for detection, diagnosis and production of
CC transgenic animals. The present sequence represents the amino acid
CC sequence of KIAA380.

XX Sequence 1520 AA:

Query Match 81.9%; Score 6455.5; DB 20; Length 1520;

Best Local Similarity 82.2%; Pred. No. 0;

Matches 1276; Conservative 89; Mismatches 129; Indels 59; Gaps 12;

QY 1 MSIRLPHSIDNSAKQSHSSPLASWLSUSLSDSTPPTSPSHRQPSDSTETAGL 60

Db 1 msvirpshidnsakqshssplawlsuslsgdstpptspsrhqpsdsdettgl 43

QY 61 VORCVTIQKDHGEGFTVSGDRIYVSVPRGAAKMGKVGEGDIIKVNGLMVTNNSHL 120

Db 44 vgrcvliqkdhngftvsgdriyvsprgaaamkagvgegdlikvngltmvtvnsahl 103

QY 121 EVVKLIKSGAVALLLGSSPPSVSGIQONPSVAGVLRNPITPPPPPPPLPPPHI 180

Db 104 evvklksgavallllgsspsmgisglqdpagapritsvlpsppppplpppprfl 163

QY 181 TGPKPLDPEYOKAATQILMNLROEBELDILPCPGTSGORTCEGLSDSGADSGL 240

Db 164 tgpklpdpdyokaatqilmlnlroebelqdlplpydtsqrpsegrlsidsqgdsdl 223

QY 241 DSGTRFSPSISSESLMNRNSVSLDPLDPSPTVLARVAOHRRGSGDALPLPHOGI 300

Db 224 dsgrtfrpsisesslmnrnsvslsdpldsprtsvlarvvaqhrrtgsdaavpsltdgqv 283

QY 301 DQSPKRLIGDEEDYDPCVFNNESDIIPODLEKLSHPAYLVFLRYLISQADBPGLLEY 360

Db 284 dqspkrligdeedydpcvfnnesdiiqdlleklsrphaylvflrylfsqadbppllfy 343

QY 361 LCSEYQOOTNPKDRSLGKDIWNIFLEKNAPLRKIPMLQAEIDLRLNNEPRNVICE 420

Db 344 lcsedyqootnpkdrslgkdiwnifleknaplrkkipmlqaeidlrlnneprnvlice 403

QY 421 AOEAVMLEIOEINDYRSKRTLGSLVGENDLLGLDSDPLREBOMAEKOLAAGDITLSK 480

Db 404 ageavmleioeindyrskrtlgslvgenldllglddplrerbyaekqlaagdltsk 463

QY 481 YEEDRSAPMDRAVNTFMNHAGIRLRESRSCTAEKTSAPRDKWLPFPPTKRSNSK 540

Db 464 yeedrsapmdravntfmnhagirlresrsctaeaktosaprdkwlpfpptkrnsnsk 522

QY 541 KEKQALDEKRNPIIRYIGKPKSSQS-----IKPGVNRNIIQHFENSQYDVE 590

Db 523 kekqaldekrnpiliryigkpkssqs-----ikpgvnrniiqhfenngydepe 582

QY 591 PGTORLSTGSPFEDLIESDSSRSSEIRLGRSGSLGREEMKRSKRAENPRRSDVMDMA 650

Db 583 pgtorlstgspfedliesdssrsseirlgreslkgreemkrskraenprrsdvmdma 642

QY 651 AEAARLHQSASSSASLSTRSLLENPTPEPTKMGRRSIESINLFCCTDVILPHILEDLIG 710

Db 643 aeearlhqsasssaslstslslennptpeptkmgrrsiesinlfcctdlvlphilledlg 702

QY 711 QLSDLPEPEVONMOHTYKQVYVANLTQRETDROEVINELFVTEASHLRTLVLDLITYQ 770

Db 703 qlsdlpepevonnhtykvdyvvanltqretdroevinelvteashlrtlrvldlityq 762

QY 771 RMREKMLPRRELRLRPNLPLELIEITHNSWCEAKKLEEDPIIRSDPLAFEDGAR 830

Db 763 rmkenlmprelrlrpnlpelleithnswceakkleedpiirsdplafedgar 822

QY 831 EELQOVAOFCSSYVALLEIRTKORKESRQFLQMOEASHPOCRRLDLRLIVSEMORL 890

Db 823 eelqovaofcscyvalleirtkorkesrfqlmqeashpocrriqlrldlissemqrl 882

QY 891 TKYPLLENIITKHREGTSEHEKULCRADOCREILKFVNEAVKQTEHNHRLLEGYOKRIDA 950

Db 883 tkypillesitkhregtsehekulcradocrellkfyneavkqtehnhrlegyokrida 942

QY 951 TALERASNPAAEFKSLDLTRKMTIHGRLTMRISKDITLQVLLIEDLVLLQROBER 1010

Db 943 talerasnpaaefksldltrkmtihgrltmriskditiqlvlliedlviqlqdek 1002

QY 1011 LLKCHSKTAVGSSDSKOTFSPVILKLNANVLIRSVATDKRAEFTICTSELAPQIYELVAL 1070

Db 1003 llkchsktavgsdskotfsvpvlklnavllrsvatdkraeffictslgppqlyelval 1062

QY 1071 TSSDKNIMELLEFAVONATKHPGAAPRPHPSPPGSGEPAYOGSTSRVETINSEVYHT 1130

Db 1063 tssdknimellefavonatkhpgaaaprphpsppgsgepayogstsrvetinsedvhtg 1122

QY 1131 EKEPKKLPEGGPPEORVODKOLIAOGPEVOR--EDDEELRLPAPAPLSDENGRIRTRD 1188

Db 1123 ekepkklpegppeorvodkoliaogpevor--eddeelrlpapaplsdengrirtrd 1182

QY 1189 PVLALATGLPLMEGLADALDENVENLRHLILMSLLPGHTVTKQAAGEPEDDLTPPSVVS 1248

Db 1183 pvlalatglplmegladaldevlenlrhlilmsllpghtvktqaagepeddltppsvvs 1241

QY 1249 ITSHPMDPGSGOAPT--ISOSTRLARREGSGOPGEDEVAVSSLAHLPRRTSSGVMDSPE 1306

Db 1242 vtsnpmdpgsgoapt--isostrlarregsgopgedevavsslahlprrtssgvmwspe 1301

QY 1307 LDRNPAEAASSTEPAAASYKVVVRKVSILPGGCVGAAKVAGSN---AIPPSGSESELSSEVE 1363

Db 1302 ldrnlaeassteaagykvvrk-----aevagskvvpalpsesgsgseppreve 1350

QY 1364 GGAQATGNCIFYVSMAPGLDSDSTETGTPSPSOCHSLPAMPTEPQYRGVGGQCS--- 1420


```

Db      959  |cr-maasvkegstkpripdstpgegndeedpsklkeeghsvtglsqspdralgles 1017
Oy      1107  ---SOEPAYOGSTSSRVEIND-----SEVYTEKEPKKLPRG-----PCPE 1144
Db      1018  tllskpshstsgskseverdlfvaergfakqgltldctlkvegdgyajapdsnlhpse 1077
Oy      1145  OR-VOD-----KOLIAOGEPRVQED-EFELRFLPRAPSLDSENGITPRDVLAL 1194
Db      1078  ewaldalnlgllkqlllvqglgleksvgeowghfr-----yt-----as 1120
Oy      1195  TGPLMEGIADAALEDVENVLRHLILMSLLPGHTVTOAQAEPEDDLTPPSVASTSPHW 1254
Db      1121  qgp-----qtdsvlqnsenik---ayhsgeghmpfrtfgtdlatscysprtesfa--pr 1170
Oy      1255  DPGSGCAPTITSDSTRL-----ARPGS-QPEGEDVAVSSLAHLPPR-TRS 1298
Db      1171  d--svglapqdgasnllymdhmltptemptpepegiddsqehlfidareahsdenpseg 1228
Oy      1299  SGVMSPELDRN-----PAAEASTEPASVYKVRKVSLLPGGAGAKVA 1344
Db      1229  dgaunkeekdvnlrlisgnyllldgydpygesstdeevs-----sltqpmtdipavest 1283
Oy      1345  GSNA-IPDSGQSESELS-----EVEGAGATGNCFCYVSNPAGPLDSTE----- 1387
Db      1284  hqgthspqntbhdgaitspftpeflvqgrwamey-scfeiqgspscadsgsqimeylnk 1342
Oy      1388  -----PTGTPSPSQCHSLPAMPTEPQPRGVRGQCQSSLVARDVDV 1429
Db      1343  eadlehlkegggkltpffakgwldqpsqt-----stqikvraacpggdc-----rllld 1391
Oy      1430  IFHTI-----EQLTKLHRLKDMELAHRELKSLGESSGCTTPVGS 1471
Db      1392  eyrpoltswlqcgcreavt-----edlpvnhlglvksqgsgstxsqgs 1438

RESULT 4
AAB90743
ID  AAB90743 standard; Protein; 1461 AA.
XX
AC  AAB90743;
XX
DE  07-JUN-2001 (first entry)
XX
XX  Human CWA20_2 protein sequence SEQ ID 186.
XX
KW  Human; secreted protein; nutrient; cytokine modulator; proliferation;
    differential; immune system modulator; tissue growth; chemotactic;
    haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
    haematopoiesis.
XX
XX  Homo sapiens.
XX
XX  WO200119988-A1.
XX
XX  22-MAR-2001.
XX
XX  14-SEP-2000; 2000WO-US25135.
XX
XX  17-SEP-1999; 99US-0398829.
XX
XX  (GENY ) GENETICS INST INC.
XX
XX  Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
    PI Meiberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
XX
XX  WPI; 2001-244801/25.
XX
XX  DR N-PSDB; AAF98482.
XX
XX  Isolated nucleic acids encoding polypeptides, useful for modulating
    e.g. cytokine and cell proliferation/differentiation activity, the
    immune system and hematopoiesis regulating activity -
XX

```

```

PS      Disclousre; Page 505-511; 557pp; English.
XX
CC      Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC      proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC      tissue types, and may be used in the prevention, treatment and diagnosis
CC      of diseases associated with inappropriate protein expression. The
CC      polypeptides and nucleic acids may be used as nutrients or to modulate
CC      cytokine and cell proliferation/differentiation activity and may also be
CC      involved in modulation of the immune system. The cDNA sequences,
CC      proteins, their agonists and/or antagonists exhibit haematopoiesis
CC      regulating activity; tissue growth activity; activin/inhibin activity;
CC      chemotactic/chemokinetic activity; haemostatic and thrombolytic
CC      activity; receptor/ligand activity; anti-inflammatory activity;
CC      haematopoiesis activity; cadherin/tumour suppressor activity; and/or
CC      tumour inhibition activity, included in the invention are probes
CC      represented in AAF98490 - AAF98572 which are specific for the cDNA clones
CC      encoding the secreted proteins.
XX
SQ      Sequence 1461 AA;

Query Match      19.1%; Score 1506.5; DB 22; Length 1461;
Best Local Similarity 29.6%; Pred. No. 4.5e-93;
Matches 459; Conservative 243; Mismatches 494; Indels 357; Gaps 54;

Oy      161  VNPILPPPPPPPLPPPHITGPKPL--QDEPVOKHATQILIMNMLROEEELDITLPPCG 218
Db      1  msplmtshpsggaasnmerticsprvimgseenvhqnkveillrlnkqegerlql1-----q 56
Oy      219  ETSQRTCEGRSLVSDSQEADSGLDSTGERF-----DS-----ISESIMN 256
Db      57  edynltppqgrllkqlkeakkhlpqldqelskatgaqgavtvsrplgdltltvsaead 116
Oy      257  RNSVL-----SDPGLDSPQSPVILARVAQHHRGSDALLRLNHGIDQS 303
Db      117  pgdvlgrtdcssgdasrpsdnadspksgr--kerilyleenpekse-tigtdctslvgs 173
Oy      304  PK---PLIGPEEDYDGYFNESDII-----FQDEKLKSHPAYLVVLYRILSQAD 353
Db      174  pstriaphlgaedd---dfgteheqngcscqstgellksrpahlavflnhvsgfd 229
Oy      354  PGPILFYLCSEVYQOTNPKDSRLGKDIWNIFLEKNAPLRYKIPBMLQAEIDL---LRN 410
Db      230  patllcylsldykhlnsketrflflefghflinsahlkvsydpemsadlekrrpel 289
Oy      411  NEDPRNVCEAQAEMLEIQDQINDYRSKRRLGLSGXGNDLGLDGDPLREGRMAEKO 470
Db      290  eolnhhyiqmgevrvhpevqrnlkdfgrksmgll--laesellkldaeardkdritleke 347
Oy      471  LAALGDILSKYE-----EDRSAPMDFAVNTFMSHAGIRLRESRSCAETKOSAP 520
Db      348  rtcaeqlvakleevlmaxgavxekdstmgvllmymkhlgvkvxep-----nle 398
Oy      521  DKDKWLPFPPTTKROSSSKKEKDALEDKRNPIRLYRGKPKSSQS----- 567
Db      399  hkrgriflplkfq---smkdkkegeekgkrgrfslipprprphnsaigramejqk 455
Oy      568  -----IKPGVNRNIQHFENSHOYDVPBPTQSLGSPPEDELDSSRSSEIRLGRS 620
Db      456  arhpkhlistpsvsspepdasaklrgsglanegt---dasyjpa-----nmsvasasas 506
Oy      621  GSLKGREEMKRSKAENVPKPSDYDMAAEARLHOSASSASLSSTRSLNPTPPF- 679
Db      507  fsqeggke-----ndtgsqyvgetsapgdtdlgtrptlnt-vdfdfpppld 551
Oy      680  -----TRKMGRRSTESPNLGFCTDVIILPHLEDDDLGQLSDLEPPPEYQN 723
Db      552  qvgeeeceverytehnpdkpfrk-fdsvalfegsgs-----edeqfe-nletpd--pn 600
Oy      724  WQHTGKDVVANLNGRELROEVINELFVTESHSRLRTLVLDLFYQWRKRNLMPPREL 783
Db      601  wqqlsvrevlllgkpkceikrgevlnellfytetanvrltkvidqvlyqvtrvsgllspeel 660

```



```

Db      313 tpgqd-----npyvslhplstcdsvdsrep-gvdtcpqpgddppqpcslap----- 357
Oy      699 VILPHLEDDLGOLS-----DLEPPEVONQNMHTYGRKOVANLQREIDR 743
Db      358 -lappedstedgetespepgddgeprsgtlepe-epppwrelvppdtllslpksqvr 415
Oy      744 QEVINELFVTEASHLRTLRVLDFYQRMKRENLMPREELRLPFLNELIHNMSCE- 802
Db      416 gevlsellvteahvrmrlrvhldlrypmadggffrldeqlnlfpsldellelvhsfldr 475
Oy      803 AMKRIREGPIIRDISDPLARFDPGARELEQOVAQFCYSQVALERIKQKRESRFQ 862
Db      476 lmktrgesgylleeigdvllarfdaegswfqlksrfsrqsfaledlkakqkxepfc 535
Oy      863 LFMEAESHPQCRRLQDLVSEMQRLTKYPLLENIKHTEGGTSHHEKLCARQDCR 922
Db      536 atvgeaesrptcrllqldmrlptemqrllkypdllqslqgnleest-ergkveaaecor 594
Oy      923 ELIKFVNEAVKOTENRHRLEGYOKRLDATALERASNPAAEFKSIDLTRKMIHEGPLTW 982
Db      595 ellhvnqavrdmedllrldkygrrldlthlrgsdpmlsefkdldtkkklvnegrltv 654
Oy      983 RISKRKITDLOVLLLEDVVLQREQEKLKCHSKTAVGSSDSKQRTSPVLKLANAVLIR 1042
Db      655 rvtckakalevhlldldlllllqrgderllkshsrlltpdpdktmlrvlrltsamtr 714
Oy      1043 SVATDKRAFFITCTSELPPDIYELVALTSSDKNIMWELLEEAQONATKHPGAPIPPIH 1102
Db      715 evatdhkafyrlftwd-geaqlyelvaqtsserknwncllletqslkvppasarlkprp 773
Oy      1103 SPPGSGEPAYOGSTSSRVEINDESVYHTEKEPKLPES-----PGPE-----GRVQ 1148
Db      774 spsrlrpepll---sssengtgaemapadarterllndllpfcrrppegqlaatalqvl 830
Oy      1149 D-KOLIAAGEVQEDDEELTLPRAPSLDGEN-RGIRTRDPV-LLALTPRLMEGLAD 1205
Db      831 slkqdlis-----teeds-----gaqpprtdgdyvpgtrpapyhltgeellslsleval 879
Oy      1206 AALDVE---NLRLILMSLLPGHTVTKQAAGE 1235
Db      880 rqlleleeeefcrlrpll--sqllgltlspnlape 911

RESULT 6
AAW42103
ID AAW42103 standard; Protein; 912 AA.
XX
AC AAW42103;
XX
DT 14-SEP-1998 (first entry)
XX
DE Human p115 Rho-guanine exchange factor.
XX
KW Guanine exchange factor; p115 Rho-GEF; human; GTPase; cancer;
KW metastasis; apoptosis; bone resorption; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN MO9820127-A1.
XX
PD 14-MAY-1998.
XX
PF 07-OCT-1997; 97MO-US15851.
XX
PR 06-NOV-1996; 96US-0029979.
XX
PA (ONXY-) ONXY PHARM INC.
XX
PI Hart MJ;
XX
WPI; 1998-286943/25.
DR N-PSDB; AAV32024.

```

```

XX      115Rho-guanine nucleotide exchange factor - useful for diagnosis
PT      and treatment of diseases involving GTPase, e.g. cancer, bone
PT      resorption etc.
XX
PS      Claim 4; Fig 1; 66pp; English.
XX
CC      This is the deduced amino acid sequence of novel human p115
CC      p115 Rho-guanine exchange factor (GEF), a polypeptide that has a
CC      specific binding affinity for a guanine nucleotide-depleted state
CC      of G-proteins (in particular Rho), a guanine nucleotide exchange
CC      activity, an oncogenic transforming activity, and an immunogenic
CC      activity. GEFs modulate cell signalling pathways, both in vivo and
CC      in vitro, by modulating the activity of GTPase. The p115 amino
CC      acid sequence was deduced from a human foetal brain cDNA sequence
CC      (see AAV32024). p115 Rho-GEF and its fragments can be expressed in
CC      recombinant host cells, and used to modulate the activity of Rho or
CC      more generally any GTPase, as well as to screen for (ant)agonists
CC      able to treat diseases or conditions involving GTPase. p115 can be
CC      used to raise antibodies useful for detecting p115 for diagnosis
CC      and research, while fragments of p115 with (ant)agonist activity
CC      are used to treat diseases/conditions where a GTPase is involved,
CC      e.g. cell proliferation (cancer), growth control, morphogenesis,
CC      stress fibre formation, embryonic development, metastases,
CC      apoptosis and bone resorption. Interaction of p115 with Rho
CC      promotes cell proliferation and/or induces cytoskeletal
CC      rearrangements.
XX
SQ      Sequence 912 AA;

Query Match      15.2%; Score 1196.5; DB 19; Length 912;
Best Local Similarity 34.2%; Pred. No.2.5e-72;
Matches 335; Conservative 167; Mismatches 324; Indels 153; Gaps 32;

Oy      308 ITGPE-EDYDPGTYNN--ESDITQDLKLSHPAYLVLFRLYLSQADPEPLFYCSE 364
Db      23 llgaeddfeneletnseeqnsqfsgleykvrrpahalmlqhvialqfepyllcllnad 82
Oy      365 VYQQTNPKDSRSLGKDIWNIPLFLEKNAPLRVKIRPMLQAEID---LRLRNNDPPNVCFA 421
Db      83 mlgslpghkaakafldlyhsflektavlrpvpunvafeldtraddlasedvqrrtvev 142
Oy      422 QEAVMLEIOEQINDYRSKRITGLSLYGENDLGID-----GDPRLRROMAEKOLAA 473
Db      143 vsgsqvavvgrqldefrkrimgmtr--wegelaglaevwvrrdrasysaerthvaerllmh 200
Oy      474 LGDI--LSKYEEDRSAPMDFAVNTFMSHAGIRLRESRSCTAEKTOGAPDKDKLPPF-- 529
Db      201 leemghlstddeeksaavnaigymthlgyrtksq-----dkksgrnfrfk 247
Oy      530 -----PKTKKSSNSKKEKDLAEKDKRPIRLYIGKPKSSOSIKPGNVRNIIQ 578
Db      248 kvmgnrtsdpprpktkglssi---ldaarwnrgeqyvpdfrlhkaevdaekpg----- 297
Oy      579 HFENSHQDVPEPCTORLSTGSPEDLLESDDSSREIRLGRSGSLKGREMKRSRAENV 638
Db      298 -----atdkgygvmp-----std-----rnlgapqgdclpysvllnpisl 331
Oy      639 PPRPSDVMDAAAEARLHOSASSASSLSTRLENPTPPFKMGRRSIESPNLGFCTD 698
Db      332 dsprdrepadaplel-----gdsapqgmstleslappedstedget-esppeg----- 378
Oy      699 VILPHLEDDLGQ--LSDLEPER-EVQNMQHTYGRKOVANLQREIDROVINELEYTEA 755
Db      379 -----degeprysgylelepeepgywrelvppdtllhlpksqykrgevisellvte 429
Oy      756 SHLRTLRVLDFYQRMKRENLMPREELRLPFLNELIHNMSCE-AKKXIREGPII 814
Db      430 ahvrmrlrvhldlrypmadggffrldeqlnlfpsldellelvhsfldrlmktrgesgyl 489
Oy      815 RDISDPLARFDPGARELEQOVAQFCYSQVALERIKQKRESRFQLEMQEAESHPOC 874

```

```

Db 490 eelgvdllarfdgaegswfqtksrfsrsgfaleqlkagqrkdpfcfaivgeaesrpic 549
Qy 875 RRLQLDLIVSEMQRLTKYPLLENTIKHTEGTSSEHEKLCARADOCREILTKFVNAVKQ 934
Db 550 rrlqldmptemqrltkyp1llqsgnteept-erekvelaacecrellhvnqavd 608
Qy 935 TENRHRLEGYOKRLDTALERASNP1AAEFKSLDTTRKMIHEGPTWRISKDTLDLOV 994
Db 609 medllrlkdygrldshlrqssdpm1sefkndltckk1vhegplrtvvtckavevny 668
Qy 995 LLELEDVLLROBERLLKCHSKTAVGSSDSKQTFSPVYLKNAVLIRSVATDKRAFIIT 1054
Db 669 lllld1lllllqgderllllkshsrcltppdgktmlrplrltsamtrevatkafayl 728
Qy 1055 CTSELGPPQIYELVALTSSDKNIMWELLEVAVON-ATKHGAP1PIHPSPOSPAYO 1113
Db 729 ftwd-geaqlvelvaqveerkmwcalitetagslkvpasprpk-rprpstrelp1ls 786
Qy 1114 GSTSSRVEIINDSEVYHTEKEPKKLPG------PGEORVOD-----KOLIAQGP 1158
Db 787 sseagn----gretspadarterlledllpfcrrpgeq1aatalkvls1kqllf---p 840
Qy 1159 VOEDEDEELRTLPRAPSIDGEMRGIRTRDPVLLATGPILMGLAD-----ALEDVE- 1212
Db 841 aeedn------gagprp1dgd--gvpg99gplspartge-igenllsleemkqleee 889
Qy 1213 ---NLRHILWLSLLPGHTV 1228
Db 890 eefrlrpl1--sq1ggnsv 906

RESULT 7
AAV41009 standard; Protein; 912 AA.
AC AAV41009;
XX 06-DEC-1999 (first entry)
DE Amino acid sequence of p115 Rho GEF.
KM RGS; guanine nucleotide exchange factor; GEF; RGS domain; GAS; cancer;
KM G protein alpha subunit; cell proliferation; growth control; hemostasis;
KM morphogenesis; stress fiber formation; integrin-mediated interaction;
KM embryonic development; tumor cell growth; cell death; leukocyte homing;
KM bone resorption; clot retraction; db1 homology domain; mechanical stress;
KM plectstrin homology domain; p115 Rho GEF.
OS Homo sapiens.
XX
XX WO9947557-A2.
XX
XX 23-SEP-1999.
XX
XX 18-MAR-1999; 99WO-US06051.
XX
XX 18-MAR-1998; 98US-0078634.
XX
XX (ONYX-) ONYX PHARM INC.
XX
XX Bolland G, Hart MJ, Roscoe W, Polakis P, Sternweis P, Kozasa T;
XX Jiang X;
XX
XX WPI; 1999-571821/48.
XX
XX N-PSDB; AA22281.
XX
XX New isolated RGS-GEF polypeptides, used to develop products for
XX modulating, e.g. cell proliferation and integrin-mediated interactions
XX
XX Examples; Fig 10; 75pp; English.
XX
XX The invention relates to isolated RGS-guanine nucleotide exchange factor

```

```

CC (GEF) polypeptides. The novel isolated RGS-GEF polypeptide consists of an
CC RGS domain of a GEF protein and does not comprise a db1 homology (DH)
CC domain or a plectstrin homology (PH) domain. The RGS-GEF polypeptides can
CC be used for modulating an activity of a G protein alpha subunit (GAS).
CC The products can be used for the regulation of biological pathways in
CC which a RGS-GEF polypeptide is involved, particularly pathological
CC conditions, e.g. cell proliferation (e.g. cancer), morphogenesis, growth
CC control, stress fiber formation, and integrin-mediated interactions, such
CC as embryonic development, tumor cell growth and metastasis, programmed
CC cell death, hemostasis, leukocyte homing and activation, bone resorption,
CC clot retraction, and the response of cells to mechanical stress. The
CC products can also be used for detection, diagnosis and production of
CC transgenic animals. The present sequence represents the amino acid
CC sequence of p115 Rho GEF.
CC
XX
SQ Sequence 912 AA:

Query Match 15.2%; Score 1194.5; DB 20; Length 912;
Best Local Similarity 34.4%; Pred. No. 3,4e-72;
Matches 339; Conservative 165; Mismatches 316; Indels 165; Gaps 33;

Qy 308 ITGPE-EDYDPCGFNN--ESDIIIFODLEKLSHAYLVFLRYILSOADPGPLFLGSE 364
Db 23 llaeededfeneletcseeqnsqfqlsqgvrkrpanhmal1qnvatqfeppllcclhad 82
Qy 365 VYQQTNPKDSRLGKDINWIFLEKNAPLRVKIPEMLDAEID--LRLNNEDEPRNVLCBA 421
Db 83 m1gslgpkaeakkafldfynshfektavrlvvpvnafe1dtradd1sedvgrtfrvge 142
Qy 422 QEAVMLEIOEQINDYRSKRTLGSLYGENDLGLD-----GDPLEROMAEKOLAA 473
Db 143 vsqgvaavgrqledftrskrlmgmcp--wegelaqleawgrdrasyearvaer1vaelmh 200
Qy 474 LGDI--LSKYEEDRSAPMDFAVNTFMSHAGIRLRESRSSCTAEKTOASAPDKKMLPF- 529
Db 201 leemqhtlstdeeksaavna1gylmrhlygrtksg-----dkksgrnffik 247
Qy 530 -----PKTKQSSNSKKEKDALDEKRNPLRTIKGPKSSSOSIKGNVNRNTIQ 578
Db 248 kymgnrrsdppkktkq1ss1--l1daarwnrgepvpdfrl1kave1d1kpg----- 297
Qy 579 HFNESHQYVPEPGQRLSTGSFPEDLSDSSRSIRLGRSGSLKRGEMKRSKAENV 638
Db 298 -----atdrkgygvmp-----s1d-----rn1gapdgqtpgvs1hps1 331
Qy 639 PRFRSDVMDAAEAARLHQASSSASLS1STRLENPRPFPFKMGGRSIESPN1GFC1D 698
Db 332 dsprdregeadap1el-----gdsppgqpm1se1lappestdegaet-eespepg----- 378
Qy 699 VILPHLEDD1CQ--LSDEPER-EYQNMQHTVAGKDVANLQREIDREGVNELFVTYA 755
Db 379 -----degeprgsq1elepeepgvr1elpdt1h1pksqvk1geve1lvt1ea 429
Qy 756 SHLRTLVLDLIFYOMRKENLMPRELARLPN1PELEINNSWCE-AMKRLREGPIT 814
Db 430 ahvmlrv1h1d1f1qpmace1ff1ee1qn1f1ps1de1lev1s1f1dr1lmk1rrgeeg1y1 489
Qy 815 RDISPMLARFGDPAAREELQOVAOFCYSQSYVALDELIRTKORRESFQLEFMQEAESHPOC 874
Db 490 eelgvdllarfdgaegswfqtksrfsrsgfaleqlkagqrkdpfcfaivgeaesrpic 549
Qy 875 RRLQLDLIVSEMQRLTKYPLLENTIKHTEGTSSEHEKLCARADOCREILTKFVNAVKQ 934
Db 550 rrlqldmptemqrltkyp1llqsgnteept-erekvelaacecrellhvnqavd 608
Qy 935 TENRHRLEGYOKRLDTALERASNP1AAEFKSLDTTRKMIHEGPTWRISKDTLDLOV 994
Db 609 medllrlkdygrldshlrqssdpm1sefkndltckk1vhegplrtvvtckavevny 668
Qy 995 LLELEDVLLROBERLLKCHSKTAVGSSDSKQTFSPVYLKNAVLIRSVATDKRAFIIT 1054
Db 669 lllld1lllllqgderllllkshsrcltppdgktmlrplrltsamtrevatkafayl 728

```

Query Match	Best Local Similarity	Score	DB	Length	846:
Matches	297:	Conservative	142:	Mismatches	258:
				Indels	205:
				Gaps	30
QY 670	RSLENTPT-PTTPKMGKRSIESPNLGFCDVLLPHLEDDLDLSLEPEPEYOMQHTV	728			
Db 32	rvtehtgtpkf-----rkfsvatgsgs-----edeqte-ndlecdp-pmqqlv	75			
QY 729	GKDVAANTLQREIDROEVINELFTVBASHLRPLRLVLDLFTYOMRKKNIMPREELARLP	788			
Db 76	srevllglkpcelkrgvevneilfyetarahvrltkvidgfygrsvregllspseilrkifs	135			
QY 789	NLPELIEIHNSMCEAMKRLRE--EGPITRFDISDPMILARPDGARELDOOVAOFCYSQSV	846			
Db 136	nledllqhlghlneqmkavtrknevsvdqjdedlltwfsgpeekllhaaatfcsngqf	195			
QY 847	ALELRTQKRSKSRQOLFQMAESHPQCRRLDRLDLYSEMORLTYUPLLENYINKETG	906			
Db 196	alemlkrsqgkksrftqgtfvgdaesnprrcrlqldkldprgmqrlltkyprlllnhiatyew	255			
QY 907	GTSHEHLKCRARDQCRELILFVNEAVKOTENRHLRLEGYOKRLDATALERASNPMLAEKFS	966			
Db 256	pt-erekvkaadhorgtllnyvngavkaeenqgrlgedygrlditslsikiseypnveeln	314			
QY 967	LDLTRKMIHEBRLLWRISKKOTLDLQVLLLEDDLVLLDOROEERLLKHSNTAVGSSDS	1026			
Db 315	ldltrkmlhegrlwlkvwkrtdkldlylltllledllyllqgddrllvrfhsxllastads	374			
QY 1027	KOTSPRYKLNLVNLRSVATDKRAFFICTSELSGPOIYELVALTSSDKNIMMELLEAV	1086			
Db 375	khtfsprvklslcvlyrgyatkakalfvmsmdng-aqlyelvaqvsvsektwgdqllcr-m	432			
QY 1087	QMATKHPGAAPIRHPSPG-----	1107			
Db 433	aaavkegvtkprlpqstprggedndeedpsklkeeqhjsvsglqspddlgleatllss	492			
QY 1108	QEPAVQGSTSSAVELND-----SEVYHTEKEPKKILPEG-----PEPOR-VQ	1148			
Db 493	kpsjstlsgtsqsevrddlfvaerqgafakehntqgtllkevevdyqiaipdsllvseerval	552			

QY 1304 SPELDNR-----PAAEASTEPASYYKVVKYSLLPGGCGVGAAYAGSNA- 1348
 Db 704 keekdvnlrnsnyllldgydpvgsstdeevas-----slltqgmtdipavesthqqh 758
 QY 1349 IPDSGQSSELS-----EVEGAQATGNCFFVSMAGPLDSTETPTGTPPSPSOCHS 1400
 Db 759 spqnlhdsqaslpftpeflvqrgwamey-scfeiqspsscadsgsq----- 804
 QY 1401 LPAMPTPEQPRGVGGCCSLVRDNDVIFHTTEOLTIKLRKMDLAHRELKLSIGG 1460
 Db 805 -----imeyhlkicadlenhkkveesytlilcqr1ag 835
 QY 1461 ES 1462
 Db 836 sa 837

RESULT 10
 AAM41561
 ID AAM41561 standard; Protein: 846 AA.
 XX AAM41561;
 AC
 XX 22-OCT-2001 (first entry)
 DT
 XX Human polypeptide SEQ ID NO 6492.
 DE
 XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 PN WO200153312-A1.
 PD 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 XX WPI: 2001-442253/47.
 DR N-PSDB; AAI60717.
 DR
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2; SEQ ID NO 6492; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SO Sequence 846 AA;

Query Match 14.0%; Score 1100.5; DB 22; Length 846;
 Best Local Similarity 32.9%; Pred. No. 7.4e-66;
 Matches 297; Conservative 142; Mismatches 258; Indels 205; Gaps 30;

QY 670 RSLKNPT-PTPKMGRSISPNLIGCTDYILPHLEDDLGQSLDEPEEYONMOTV 728
 Db 32 rvtetgtpkpf-----fkdsvalfgesqs-----edeqfe-ndletdp--pmwqlv 75
 QY 729 GKDVAANLTGREIDROEVINELFTVTEASHLRLRYLDIFYQRMKKNLMPREELARLP 788
 Db 76 srevllgkpcelkrqevlnelftyerahvrlkvldgvlyrsreglspselrkifs 135
 QY 789 NLPELIEIHNSWCAMKKLRE--EGPITRDISDPMARFDPAREELOQVAQFCSYQSV 846
 Db 136 nledlqlghlneqmkvkrknetsevidqgedliltfsgpgeeklkhaattfscngpf 195
 QY 847 ALELRTRKRESRQLEFMOEAESHPOCRQLQDLIVSEMRQTLKYPLLLENITKITEG 906
 Db 196 alemksrqkdsrftqvtvgaesnprrlqkldlptqmrlltkypyllldnatylew 255
 QY 907 GTSHEKICRARDOCRELKFVNEAVKOTENRHNLEGGQKRLDNTALERAANPLAEKRS 966
 Db 256 pt-erekvkaadrcqllnyngvkaenkqrledyqrldsslkiseypvveeln 314
 QY 967 LDLTRKMINHGRLTWRSKTKTLDLYLLEDDLVLLROEROERLLKCHSKTAVGSSDS 1026
 Db 315 ldlckrmlhegplvwkvrcktdllylllledlilvlkqgddrlvlrckskl1astads 374
 QY 1027 KOTSPVILKNAVLIRSATTKRAFFITCTSELPQIYEVALTSSKKNIMLEDEAV 1086
 Db 375 khtispylklstvlrvqatnkalfivsmesdg-aqlyelvaqtveksltwqlllcr-m 432
 QY 1087 QNATKHPGAAPLPIHPSPG-----s 1107
 Db 433 aasvkegstckpdlpqscpgegdndeepsklkeeqngisvtqjgsprdlglestliss 492
 QY 1108 QEPAYOGSTSSRVEIND-----SEVYHTEKEPKKLPPEG-----PGPEOR-VQ 1148
 Db 493 kpgshlsatgsksevrldfvaerqfakeqhtdgtlkevgedyqlaipshlpvseerwa1 552
 QY 1149 D-----KQIADGEPVQED--EELRLTPRAPPSLDGNGRIGTRDPVLLATGRL 1199
 Db 553 dalrnlglkqlvqqlgltetksvqedwqhfr-----yrt-----asqg- 593
 QY 1200 MEGGLADALEVDENRHLILMSLRLPGHTVKTOAGGEPEDDLTPRPVSYSIHMPDPESP 1259
 Db 594 ---qtdsvlqnsenik---ayhsgegimpfritgtdatcysprstestfa--prd--sv 643
 QY 1260 GQAPRTISDTRL-----ARPEGG-OPEGEDVAVSSLAHLPPR-TRSSGWD 1303
 Db 644 glapqdsqasnllymdhmtprempmepegldsgchffidareahsdempsegdga 703
 QY 1304 SPELDNR-----PAAEASTEPASYYKVVKYSLLPGGCGVGAAYAGSNA- 1348
 Db 704 keekdvnlrnsnyllldgydpvgsstdeevas-----slltqgmtdipavesthqqh 758
 QY 1349 IPDSGQSSELS-----EVEGAQATGNCFFVSMAGPLDSTETPTGTPPSPSOCHS 1400
 Db 759 spqnlhdsqaslpftpeflvqrgwamey-scfeiqspsscadsgsq----- 804
 QY 1401 LPAMPTPEQPRGVGGCCSLVRDNDVIFHTTEOLTIKLRKMDLAHRELKLSIGG 1460

DB 805 -----Imeylhkheadlehlkveesylllqrlag 835

QY 1461 ES 1462

DB 836 sa 837

RESULT 11

ABG18434

ID ABG18434 standard; Protein; 848 AA.

AC ABG18434;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #18425.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS

PN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Dmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX N-PSDB; AAS82621.

DR

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 48793; 103pp; English.

PS

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers, and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SO Sequence 848 AA:

Query Match 13.5%; Score 1068; DB 22; Length 848;

Best Local Similarity 32.3%; Pred. No. 1.2e-63;

Matches 281; Conservative 137; Mismatches 239; Indels 212; Gaps 25;

QY 713 SDLEPEVOMQHTGVKDVANLTQREIDROEINELFVTEASHRLRTLEVLDFYQRM 772

DB 54 ndletdp--pnwqqlvsrevllglkpcelkrgevinelfyterahvrtlkvlidvfyqr 111

QY 773 RKENLMPREELANLFNLPPELIEIHNWCCEAMKLE--EGPIIRDISDPMARFGPAR 830

DB 112 sregllspselelrfkifsnldqlqihglnegmkavkrnetsvldlgedllwfsapge 171

QY 831 EELQOVAACFQSVQVALELIRTKORKESEFOLFMEQAEASHPCRRLOLRDLVSEMR 890

DB 172 ekllhaaafcsnqfaleniakrgkdsrfgfvdgaesnpclrlqldlptcmql 231

QY 891 TKYPLLENIIRKTEGTSHEKLCRAPDQCEELFVNEAVKQTENRRL--EGYQKRLD 949

DB 232 tkypyllldlakylewpt-erekvkaadhrcqlnlfvngavkeankqlleedygrld 290

QY 950 ATALERASNPAAEFSSDLITRKMTHEGFLTKRISKDITLDQVLLLEDLVVLOROE 1009

DB 291 tselklseypuverlrnmldtkrkmlhegplvwkvnrdrkldllyllledllvllqkqdd 350

QY 1010 RLILKCHSKTAVGSSDSKOTFSPVLKLNANVLIRSVATDKRAFFICTSETLPPQIYELVA 1069

DB 351 rlvtrchsklstaadskntfsvklstvlvqvadtkalfvismdng-ajlyelva 409

QY 1070 LTVSSDRNIMWELLEAVONATKHPGAAPIRHPSPGSEPAVQGSTSRKEINDESVYH 1129

DB 410 qtvsekltwqdlfcr--maasvkqscfkprlpqstpgqd-----nd----- 450

QY 1130 TEKEPKKLPGE-----PGPEQRVODKQLAAGEP-----VQEDDEELRTLPRAPPS 1176

DB 451 -eepesklkeeghglsvtlqspdrldglsclsskpsghsistgsksevalfvaerq 509

QY 1177 LD-----GENRGIRTRDPVLLALVTGPLMEGLADALEDVENLRHLIT----- 1219

DB 510 fakeqtdgtlkvegedyjalpdsnl-----pvskerraldelnrlgllkqlvngql 564

QY 1220 -----WSLPGHTVKQAAGEPEDDLTPTPSVVSTISHPWPDSGPGCAPTISDTRL 1271

DB 565 teksalenwqhfprryafqg---pqld-----sviq----- 593

QY 1272 AREPGSQPEGEDVAV--SSLAHLPRTRSSGVWDSPELDNRNPAEAASTEPAASYVVRK 1329

DB 594 -----nsenlkayhsgeqimprftgcgdl-----atcysprstse----- 628

QY 1330 VSLPFGGVGAAKVAGSNAIIPDSGQS-----ESELSEVEGGAOATGNCFYVS 1376

DB 629 -sfaprdsvgl-----pqdsqasnllvmdhmlmtpepmepgeglldsgenhfda 679

QY 1377 MPAGPLDSSTPTGTPSPSPQCHSLPAMPTEPOPYRNGVGGQCSLVRRDVDVIFHTIQ 1436

DB 680 rhaahsdpsqdg-----avnkkeadv----- 702

QY 1437 LTIKHLRLKDMELAHRELKLSLGESSGCTTPVSGHTEAARMTDVSLSPPAKKALASPS 1496

DB 703 -----nlrtisgnyllld-----gydprgeestdeavasslsqpmtylpaves 746

QY 1497 ONGOEGSCPEGSIDALDSATIDTAVSP 1525

DB 747 -----plqtatfslrnthsdgelsp 766

RESULT 12

AAV41012

ID AAV41012 standard; Protein; 2559 AA.

XX

AC AAV41012;

DT 06-DEC-1999 (first entry)

XX

XX Amino acid sequence of DRHOGEPZ.

KW RGS; guanine nucleotide exchange factor; GEF; RGS domain; GAs; cancer;
 KW G protein alpha subunit; cell proliferation; growth control; hemostasis;
 KW morphogenesis; stress fiber formation; integrin-mediated interaction;
 KW embryonic development; tumor cell growth; cell death; leukocyte homing;
 KW bone resorption; clot retraction; cell homology domain; mechanical stress;
 KW pleckstrin homology domain; DrhGEF2.
 XX
 OS Drosophila sp.
 PN MO9947557-A2.
 XX
 PD 23-SEP-1999.
 XX
 PE 18-MAR-1999; 99WO-US06051.
 XX
 PR 18-MAR-1998; 98US-0078634.
 XX
 PA (ONYX-) ONYX PHARM INC.
 PI Bollaig G, Hart MJ, Roscoe W, Polakis P, Sternwels P, Kozasa T;
 PI Jiang X;
 XX
 DR WPI: 1999-571821/48.
 DR N-PSDB: AAZ22284.
 XX
 PT New isolated RGS-GEF polypeptides, used to develop products for
 PT modulating, e.g. cell proliferation and integrin-mediated interactions
 PT
 PS Examples: Fig 16; 75pp; English.
 XX
 CC The invention relates to isolated RGS-guanine nucleotide exchange factor
 CC (GEF) polypeptides. The novel isolated RGS-GEF polypeptide consists of an
 CC RGS domain of a GEF protein and does not comprise a db1 homology (DH)
 CC domain or a pleckstrin homology (PH) domain. The RGS-GEF polypeptides can
 CC be used for modulating an activity of a G protein alpha subunit (GAs).
 CC The products can be used for the regulation of biological pathways in
 CC which a RGS-GEF polypeptide is involved, particularly pathological
 CC conditions, e.g. cell proliferation (e.g. cancer), morphogenesis, growth
 CC control, stress fiber formation, and integrin-mediated interactions, such
 CC as embryonic development, tumor cell growth and metastasis, programmed
 CC cell death, hemostasis, leukocyte homing and activation, bone resorption,
 CC clot retraction, and the response of cells to mechanical stress. The
 CC products can also be used for detection, diagnosis and production of
 CC transgenic animals. The present sequence represents the amino acid
 CC sequence of DrhGEF2.
 CC
 XX
 SO Sequence 2559 AA;
 Query Match 12.0%; Score 945.5; DB 20; Length 2559;
 Best Local Similarity 20.9%; Pred. No. 1.2e-54;
 Matches 482; Conservative 280; Mismatches 650; Indels 895; Gaps 82;

OY 171 ----- 170
 Db 504 lslgtrkntekdltspsflltdflgqgrmsbqaesgmhqbhstpsqgfnpqh 563
 OY 171 -----PPLP--PQ-----HTG 182
 Db 564 qghrfketgtskgnkflslrsllfeedvppplpqrrppqlnldknagsppgsnlva 623
 OY 183 P-----KPLQDPEVQKHATQILNMN 202
 Db 624 pvsldratatpqlnrsgqqqlprstcdnspsnakskrikalkalsdp--kmsqmlilgm 680
 OY 203 ----- 202
 Db 681 esasaagaagslewdgppplpprlpgmmedmsrgscqnlagnsvgtfnyplvstc 740
 OY 203 -----LROEDELQDIL-----PPCGETSQR----- 223
 Db 741 tavqndlniafplsqrplvqqlqyqgqghqmsggatgalqtpnlgknkhrvgs 800
 OY 224 -----TCGRSLSDVDSQEADSGLDGTERFESISLM-----NRNS 259
 Db 801 spdmhprhprlckltsgswelvkdgess-ppqtpplpylssshmtvledpneumrga 859
 OY 260 VLSDPG--LDSPQ-----TSPVILARVAQH-HRROGSDAALLPLNHQIDOSPRLTI 309
 Db 860 aaagpvgvfieshqftfpmagaaaspilshshmaagand-----tqkeils 906
 OY 310 GPEDDYD-----PGYFNESDIIIFQDLERKLKSHPAVYVFLRYLSQADPGPLLEYL 361
 Db 907 medensdidepfidengfnlnlrlleae-----nvflaiflfnvysnsgpllfyl 960
 OY 362 CSEVYQGNPKDSSLSGKDINIFLEKNAPLR-VKIPMLQAEIDLRLRNEDP-----RN 416
 Db 961 itelyegtskdmckwayelnsstlvpaplswyqdeslarevduvqlqeydkveillrt 1020
 OY 417 VLCAQAEVMEIIOEINDYRSKRTLGSLGSEYEND--LLGLDDPDLEROMARKOL-- 471
 Db 1021 vflsrkrakallseqleqfegkrtaqgtlygptddklaekdklire-qldkylmpn 1079
 OY 472 -----AALGDIILSKYEEDRSAPMDFA--VNTFMSH----- 499
 Db 1080 lhaliedengspdevrkvatlsalstviylfnlrrppssivervhfvrdfskeri 1139
 OY 500 -----AGIR-----LRE-----SRSCTAERKQS 518
 Db 1140 mgknrkmmvgrhplvlrlygevtchenhcgllwgvsppgyhctdcklnlhrqcskvdes 1199
 OY 519 APDKDKWLPFPFKTKRKOSNSKKEK-----DALEDKR----- 551
 Db 1200 cpgp-----lpgakriahnakiskfmgkltprtsdvigneakrtsgdeidveltpdqt 1253
 OY 552 -----NPLIRYIGKPKSSQSIRKGNVRNIIOHFNSHQYDPERG----- 592
 Db 1254 asivrqpsdrpdanlsirsgntscstgl-----ntdlqsgfngscandsinp9gagc 1310
 OY 593 TORLST-----GSFPR-----DLLESSSR-----SEIRLGRS 620
 Db 1311 nmdlstvasstipstsgsvaaqlsafaelnaldetvdekarrearysqphbsapavsvrs 1370
 OY 621 GSLGREEMKRSRAEN-----VPRRSPDVDDAAEAARLHQSSASSSASSLS----- 668
 Db 1371 esykerlsnktrnrsirktspslsrip-ndeqldlqsmatlygssnssissaggsesp 1429
 OY 669 TRSLENPTRP-----FTPKMGRSIESPNL-----GFCYDYLPHLLLEDLIGOLS-- 713
 Db 1430 stmehfaaagaaagvypnqplnqnqphlilqhaqyqgqsfaglaagaasaaas 1489
 OY 714 -----DLPEPEVQ--NMQHVGVGDVYANLQRELDROEVNLEFVT 753
 Db 1490 nssfwaghpilparwtleseddvneadwsmaaveaialtaeakkrqellineiyqt 1549

```

QY 754 EASHLTFRVLDLIFQRRKKNLMPRELARLF-PNLPELTIHNSWCERAKKLK-EEG 811
DB 1550 ernhvtlklldrlffllpyesglisqdhlllfpallslreihqafegslkxrltlem 1609
QY 812 PIRIDSDPLARFDGPAREILOOVAOFCSYVALLELIRTKORESRQRLMOEAESH 871
DB 1610 hvnttgdlladmfqgsgyvvlcefaqfcarqialaeketrnkdemlqklkksesh 1669
QY 872 POCRLQRLDLIVSEMOURLTKYPLLENIKNH-EGGTSEHEKLCARADOCREILKE 927
DB 1670 kacrrtelkdlptvqlritkypillfenlykvtvrlpentteaaiqraveskrlive 1729
QY 928 VNEAVQOTENRRHLEQYQKRLDATALERASNPAAEFKSDLTRKMIHEGPIITWISKD 987
DB 1730 vngavvtaedanklqnlgkrlrdsydk-----eefkklidlqhlhnglnt-ikkn 1781
QY 988 KTLIDLVLLEDVLVLOJOREELLK-CHSKTAVGSSDSKOTFSPLYLKNA-VLIRSYA 1045
DB 1782 psvqlnglllfemvilltqddkyylnkhltpisi-----tnkvspimsidadtllirgea 1837
QY 1046 TDKRAFEIICTSELGPPQIYELVALTSDBKNIMWELLEEAVONATKH----- 1092
DB 1838 adknsffll---kmktsqmlrlapsssecktwfkhsdvaarqsknrksaenhdtsl 1894
QY 1093 --PGAPPIRHSPPSGOE-----PAYGSTS 1118
DB 1895 sdpaalaip-hsnlkeslstdltvgplaataltlttjplamjlatvtpaptnsnvs 1953
QY 1119 -----RVEINDESEVYHTEKEPKKLPEGSGPEQROVDKOLI-AGEPYO-----EED 1163
DB 1954 sltgvqlrpprdataseadqyvntpkprsgnevrnctmslstgeplqkysaangtean 2013
QY 1164 EEEELR---TLPPRPPSLDGENR-----GIRTRDPVLLAL-----TGPILMEG-LA 1204
DB 2014 dvlrlhsgstresvrgpsgeenstygmgvgnskrdasivcsnnsntrtllmgsplv 2073
QY 1205 D-AALPDVNLRLILMISLLPGHTVKTQAAGEPD-----DLTPP-----SVGS 1249
DB 2074 dptalvstspatlaevltlpgeltrldaslrndllekqkicdlfrlprvehydqivdl 2133
QY 1250 TSHPMDPGSPGQAPRTISDTRLARPEGSQ-----PEGEDVAVSSU---AH- 1291
DB 2134 ammp-----eapksadalaalaaydqitlctkmleymhvtvpegevsastavcghne 2186
QY 1292 -----LPPTRSSGVWDSPEPLDRNPAAEAASSTEPASIKYKRVSLIPGGVGAAKY 1343
DB 2187 keklrkkvapsssfssppplppnrqhaqagqippss---rlmpklqltldidev----- 2238
QY 1344 AGSNALPDGSGQSELSSEVEGGAQATGNCFYVSMPPAGPLDSSREPGTPPSPGCHSLPA 1403
DB 2239 ----alheddddyceidelt-----lpaipskpherpt-tplap----- 2272
QY 1404 WPTPEQRYGVNRGGQSSLVRRDVIYFHTI-----EQLTIKLRH- 1444
DB 2273 fhteptsgsv----idaaskrgstadvpegllleqplegdketkgednevkvtpdkls 2228
QY 1445 ----KMEIAHRELKSLGSGSGGTTPVGSRFTEAARMTDVSLSPAKALASDSQNG 1500
DB 2329 eseneerqcvaeaitkvepaditkskneaaavdelipsqrrelktaenasksvadkchedne 2388
QY 1501 EOGSCPEGSDTALEDSANTDTAVSPG 1527
DB 2389 e---tleegv-astvdsstqstpsesp 2411

```

```

DE Novel human neoplastic disease associated polypeptide #113.
XX
KW Human; neoplastic disease associated polypeptide; cancer;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic.
XX
OS Homo sapiens.
XX
PN WO200155163-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01358.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.

```

```

RESULT 13
AAU21680
ID AAU21680 standard; Protein; 203 AA.
XX
AC AAU21680;
XX
DF 04-DEC-2001 (first entry)
XX

```

PR	14-SEP-2000	2000US-02330664
PR	14-SEP-2000	2000US-02330664
PR	14-SEP-2000	2000US-02330665
PR	14-SEP-2000	2000US-02330665
PR	21-SEP-2000	2000US-02344223
PR	21-SEP-2000	2000US-02344223
PR	21-SEP-2000	2000US-02349977
PR	21-SEP-2000	2000US-02349977
PR	26-SEP-2000	2000US-02349988
PR	26-SEP-2000	2000US-02349988
PR	27-SEP-2000	2000US-02356184
PR	27-SEP-2000	2000US-02356184
PR	29-SEP-2000	2000US-02356527
PR	29-SEP-2000	2000US-02356527
PR	29-SEP-2000	2000US-02365677
PR	29-SEP-2000	2000US-02365677
PR	29-SEP-2000	2000US-02365688
PR	29-SEP-2000	2000US-02365688
PR	29-SEP-2000	2000US-02365696
PR	29-SEP-2000	2000US-02365696
PR	02-OCT-2000	2000US-02366072
PR	02-OCT-2000	2000US-02366072
PR	02-OCT-2000	2000US-02370337
PR	02-OCT-2000	2000US-02370337
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-02370440
PR	02-OCT-2000	2000US-02370440
PR	13-OCT-2000	2000US-02393535
PR	13-OCT-2000	2000US-02393535
PR	20-OCT-2000	2000US-02399377
PR	20-OCT-2000	2000US-02399377
PR	20-OCT-2000	2000US-02412221
PR	20-OCT-2000	2000US-02412221
PR	20-OCT-2000	2000US-02411825
PR	20-OCT-2000	2000US-02411825
PR	20-OCT-2000	2000US-02418069
PR	20-OCT-2000	2000US-02418069
PR	08-NOV-2000	2000US-02446524
PR	08-NOV-2000	2000US-02446524
PR	08-NOV-2000	2000US-02466174
PR	08-NOV-2000	2000US-02466174
PR	08-NOV-2000	2000US-02466775
PR	08-NOV-2000	2000US-02466775
PR	08-NOV-2000	2000US-02466776
PR	08-NOV-2000	2000US-02466776
PR	08-NOV-2000	2000US-02466777
PR	08-NOV-2000	2000US-02466777
PR	08-NOV-2000	2000US-02466788
PR	08-NOV-2000	2000US-02466788
PR	08-NOV-2000	2000US-02466523
PR	08-NOV-2000	2000US-02466523
PR	08-NOV-2000	2000US-02465224
PR	08-NOV-2000	2000US-02465224
PR	08-NOV-2000	2000US-02465225
PR	08-NOV-2000	2000US-02465225
PR	08-NOV-2000	2000US-02465610
PR	08-NOV-2000	2000US-02465610
PR	08-NOV-2000	2000US-02465611
PR	08-NOV-2000	2000US-02465611
PR	08-NOV-2000	2000US-02466113
PR	08-NOV-2000	2000US-02466113
PR	17-NOV-2000	2000US-02492007
PR	17-NOV-2000	2000US-02492007
PR	17-NOV-2000	2000US-02492113
PR	17-NOV-2000	2000US-02492113
PR	17-NOV-2000	2000US-02492114
PR	17-NOV-2000	2000US-02492114
PR	17-NOV-2000	2000US-02492116
PR	17-NOV-2000	2000US-02492116
PR	17-NOV-2000	2000US-02493217
PR	17-NOV-2000	2000US-02493217
PR	17-NOV-2000	2000US-02493218
PR	17-NOV-2000	2000US-02493218
PR	17-NOV-2000	2000US-02493244
PR	17-NOV-2000	2000US-02493244
PR	17-NOV-2000	2000US-02492445
PR	17-NOV-2000	2000US-02492445
PR	17-NOV-2000	2000US-02492456
PR	17-NOV-2000	2000US-02492456
PR	17-NOV-2000	2000US-02492965
PR	17-NOV-2000	2000US-02492965
PR	17-NOV-2000	2000US-02493297
PR	17-NOV-2000	2000US-02493297
PR	17-NOV-2000	2000US-02493299
PR	17-NOV-2000	2000US-02493299
PR	17-NOV-2000	2000US-02493300
PR	17-NOV-2000	2000US-02493300
PR	01-DEC-2000	2000US-02503916
PR	01-DEC-2000	2000US-02503916
PR	01-DEC-2000	2000US-02503921
PR	01-DEC-2000	2000US-02503921
PR	05-DEC-2000	2000US-02510388
PR	05-DEC-2000	2000US-02510388

08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
DR WPI; 2001-465558/50.
DR N-PSDB; AAS34879.
XX
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis -
XX
XX
PS Claim 11; SEQ ID No 407; 687bp; English.
XX
XX The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides, and cDNA (AAS3767-AAS35050) and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
CC neoplastic disease associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 203 AA:

Query Match	Similarity	12.0%	Score 945;	DB 22;	Length 203;
Best Local	Similarity 93.8%;		Pred. No. 3.3e-56;		
Matches 180;	Conservative 8;		Mismatches 4;		Indels 0; Gaps 0
QY	681	PMGRRSIESPMLGCTDYLPHLLLEDDDGGOLSDLEPEEVEONMQHTGVKDVYANTLORE	740		
DB	5	pmmgrrrsiespmlgctdclphlledddgqslsdlepeedaqmqhtgvkdvaygltrge	64		
QY	741	IDROEVINLEVTESHLTRLVLDLIFYQRMKKNLMPREELARLPNLPETLIHNSW	800		
DB	65	ldrgvvinellvtteashltlrldllygrmkknlmpreearlfpnlpeliehlhnsv	124		
QY	801	CEAMKRLREEGTIRIDISPNMLRFPDGPREELOQVAAOFCISQSYALIELIRTKORRESR	860		
DB	125	ceamkrlreegtlrldisdmrlarfdgpareelqgvaaqfcisyalsalelrltkrkresr	184		
QY	861	FQLEMQAESHP	872		
DB	185	fqlfmqaeeshp	196		
RESULT 14					
XX	ID	ABB62745			
XX	AB	ABB62745 standard; Protein; 2559 AA.			
XX	AC	ABB62745;			
XX	DT	26-MAR-2002 (first entry)			
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO.15027.			

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li FWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL06848.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 15027; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB17737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 2559 AA:

Query Match 11.9%; Score 936.5; DB 22; Length 2559;
 Best Local Similarity 20.8%; Pred. No. 5e-54;
 Matches 481; Conservative 279; Mismatches 652; Indels 895; Gaps 82;

QY 17 QSHLSPIASWLSLSLSDSTPRTSPSHRQPSDSTETAGLVORC-----VII 67
 DB 204 qsdnspvliqaperslntlrslrslsghtqstetpattptslajpknfilyltlv 263
 QY 68 OKDQHFQFTVSGDRIVIVOSVPRGAAKAGVKEGDRRIKVGTMVNTNSHLEVVKIK 127
 DB 264 rksnsgymkvsqdpvfvessvpgaaelaqlvagdmllrvnghevrllekhtcvvgllk 323
 QY 128 SGAYALTTLLGSS---PSVGV-----SGIAQNPSVAGLVRY----- 161
 DB 324 astvelavkrqgklrrpsavsvtrptlrsgrtdtaastltpqpqdsikrremetykq 383
 QY 162 -----NP----- 163
 DB 384 tlgkmlgekelnlerlksqdnpsyklsaanlrklreqlhqvygaedaptvklqaaankn 443
 QY 164 --IIRP-----PP-----PP----- 170
 DB 444 talltpnqiqhlsasachnqgfhllhnhnlhnnnyppqgqastpafsliprslss 503
 QY 171 ----- 170
 DB 504 lslgttknktekdltspsfgltctdfiqgrmsghaagsmgmhqhtsptsqgfthpq 563
 QY 171 -----PPLP---PPQ-----HITG 182
 DB 564 qghrfketgtakgnkflisrslleedvpprprrpprrpnrrqlnlakngnaaspggshlva 623

QY 183 P-----KPLQDPEVOKHATQILMNN 202
 DB 624 pvsdlratrspqlnrsgqqlpssldnspnaskrsklktkalsp-----kmtgmflgm 680
 QY 203 ----- 202
 DB 681 esasaagaagslevdgppplpprlpgmmedmarsgcqnlagpnsvgtafyplvstt 740
 QY 203 -----LROEEELQDIL-----PCGERSOR----- 223
 DB 741 tavqndlnlafprlsgrpnltvgqlqyqgqqlqmsggatgaltgtpnlgknhrrvs 800
 QY 224 -----TCEGRLSVDSQESADSGLDSTEREPSISESLM-----NRNS 259
 DB 801 spdmhprhprdtlcttsqswelvxkdegs-ppgtppppylsssmvtvledpennrga 859
 QY 260 VLSDPG--LDSPQ-----TSPVILARAOH--HRROGSNALPLPHOGIDOSPRLII 309
 DB 860 aaagpvgflesbqfcpmagasaplpislnshmaaqnd-----tqkells 906
 QY 310 GPREDYD-----PGYFNESDIIFODLEKLKSHPAYLVWRYILSQADPGPLFYU 361
 DB 907 medensdldepfidengpfnlrlleae-----nvflaflnyvensdpapllfy 960
 QY 362 CSEVYQQTNPKDSRLGKDIWNIFLEKNAPLR-VKIPEMIQATIDRLRNEDP-----RN 416
 DB 961 ltelkyegskdmrkwayelhsctflvrappilswyrgdeslarevdhnlqeykvelirt 1020
 QY 417 VLCEAEAWVLEIOPQINDYRSKRTGLGSLYEND--LLGLGDPLREKOMEKOL--- 471
 DB 1021 vflsrkrakdlseqrlrefqgkrttagltgllyprtdcklaeaktldre-qldkylmpn 1079
 QY 472 -----AALDILSKYEEDRSAPMFA--VNTFMSH----- 499
 DB 1080 lhaliedengspedvkrvalcesaltvylrntlrppslaverhhtvrsdkfskrt 1139
 QY 500 -----AGIR--LRE-----SRSSCTAKTQS 518
 DB 1140 mgknrkmnvtrhplvrlryevthcnhcgllwgvspggyhctdcklnlhrqskvdes 1199
 QY 519 APDKDKWLPFFPKTKKSSSKKEK-----DALEDKRR----- 551
 DB 1200 cggp-----lpqaklalnmdkiskfmgtlrptsavigneksrredeelvtlptdrq 1253
 QY 552 -----NPILRYTKPKSSSQSIKPGVNRNIIQHFNESHQYDVPEPG----- 592
 DB 1254 asivrgpsdtrpdaaisrsgntscntegl--nttdiqsfhsgcandslppggagc 1310
 QY 593 TORLST-----GSPFE-----DLESDSR-----SEIRLGRS 620
 DB 1311 mmdlstsavstlptstgsvaaglsataelnaoldtvckeartreryshpkhkaapvsnrs 1370
 QY 621 GSLGREEMKRSRKAEN-----VPRPSDVMDMAAARLHOSASSASLS----- 668
 DB 1371 esykerlnskrnrsrictsdpslsarp-ndeqldiglnaelyvgssnslsaagseap 1429
 QY 669 TSLLENPTPP-----FTTKMGRSIESPNL-----GFCIDVILPHILEDLGLS--- 713
 DB 1430 stsmehfaapgaagvqvpmglnqnhp1lllqghaqyccqdsfqaqlaagaagssas 1489
 QY 714 -----DLEPEPEVO--NMOHTVAKDVANVANTOREIDROEYINELFVT 753
 DB 1490 nsafwnaghp1pvaarvtlesededvneadwsmaaealaaltdaekkrqgeilneygt 1549
 QY 754 EASHLRTLRVLDFYQRMKRNELMPREIARLF-PNLPELIEIHNSWCEAMKKLR-EEG 811
 DB 1550 ernhvtlklldrlflpyesgllsqdhl1llfpallslsheligaefagstlkqrrlehn 1609
 QY 812 PIIRDISPMLARFDGPABEELQOVAQFCYSQVALERIKRKORKESTFQLPMQDAESH 871
 DB 1610 hvntitgdl1ladmfdqsgvv1cfeafqcarqajalealkerkndemlqk1lkkseeh 1669

[illegible]

```
Db 240 rqvatkakalfvismdng-aqiyelvaqvsekltwqdlcr-maasvkegstxpiip 297
Qy 1102 PSPPG-----SOEPAYOGSTSSRVEI 1122
Db 298 qstprgegdndeedprklikeeqhgisvtglsqspdrdgllestlisskpgshsistsgksev 357
Qy 1123 ND-----SEVYHTEKEEPKILPEG-----PPEPQR-VOD-----KOLIAQ 1155
Db 358 rdlfvaeqrfakeqhtdgtclkevgedyqlapshlpvseerwaldalrnlqllkqllvq 417
Qy 1156 GEVQEDD-EEELRTLPRAPRPSLDGENRGIRTPDVPVLLALTGPLMEGLADALEDEVNL 1214
Db 418 qlgltksvqedwqhfpr-----yrt-----asgpr-----qtdsviqnseni 455
Qy 1215 RHLIWSLLPGHTVKTQAGEPEDDLTPTPSVVITSHPWDPSPGQAPRTISDSTRL--- 1271
Db 456 k---ayhsgeghmpfritgtgdlatcysprstestfa--prd--svglapqdsqasnllvm 508
Qy 1272 -----ARPEGGS-OPEGEDVAVSSLAHLPPR-TRSSGVWDSPELDRN----- 1310
Db 509 dhmlmtpemprmepeggladdsgheffidareahsdenpsqgdqavnkkekdvnlrtsqnyl 568
Qy 1311 -----PAAEAASTEPAAGYKVVRKVSLPGCGVGAAYAGSNA-IPDSGQSESELS--- 1360
Db 569 lldgydpvqessctdeevas-----siltqpmctgipavesthqghspqnthsgaalspft 623
Qy 1361 -----EVEGGAQATNCIFYVSMRPAPLDSSTEPTGTTPSPSQCHSLPAMPTEPQPYRGVR 1415
Db 624 peflvqqrwgamey-scfeiqspscadsqsg----- 654
Qy 1416 GGCCSSLVRRDQVVFHTTEQLTKLHLRLKDMELAHRELLKSLGGS 1462
Db 655 -----lmeylhkleadlhlkveesytlcqrlagsa 687
```

Search completed: September 18, 2002, 10:37:54
Job time: 116 sec